

Real-Time IoT-Integrated Ensemble LSTM–Bi-LSTM Framework for Greenhouse Microclimate Forecasting

Siman Emmanuel^{1*}, Agu, Edward Onyebueke², Musa Gani³, Oladunjoye John Abiodun⁴, Anagu Emmanuel John⁵

^{1,2,4,5} Department of Computer Science, Faculty of Computing and Information System, Federal University Wukari, Nigeria

³Dept of Crop Production and protection, Faculty of Agriculture, Federal University Wukari.

Received: 11/08/2026 | Revised: 08/09/2026 | Accepted: 28/09/2026 | Published: 16/10/2026 |

Abstract — Accurate prediction of greenhouse microclimates is essential for optimizing crop growth and resource utilization in precision agriculture. The research presents an IoT-integrated ensemble deep learning framework for real-time greenhouse microclimate forecasting using Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks. Environmental variables, including temperature, light intensity, soil moisture, and humidity, were acquired through IoT sensor nodes deployed in a controlled greenhouse environment. A benchmark greenhouse dataset was first used for model development and hyperparameter optimization, while an original IoT-based dataset collected during cucumber, tomato, and pepper cultivation at the Federal University Wukari, Nigeria, was used for validation and real-world deployment. The data were preprocessed through noise filtering, missing value imputation, Z-score-based outlier removal, and Min–Max normalization. A weighted ensemble strategy was then applied to combine LSTM and Bi-LSTM predictions for improved robustness. Experimental results show that the Bi-LSTM model outperformed the LSTM model, achieving $R^2 = 0.9981$, $MAE = 0.0088$, $RMSE = 0.0110$, and $MSE = 0.0001$, compared to LSTM results of $R^2 = 0.9956$, $MAE = 0.0133$, $RMSE = 0.0167$, and $MSE = 0.0003$. Statistical validation using paired t-test, Wilcoxon signed-rank test, confidence intervals, and multi-run analysis confirmed that the performance improvement is statistically significant and robust. The proposed framework enables accurate real-time microclimate forecasting and intelligent environmental monitoring, providing a scalable solution for precision greenhouse management, improved crop productivity, optimized energy utilization, and efficient irrigation scheduling.

Keywords: Internet of Things (IoT); Ensemble Deep Learning; Greenhouse Microclimate Forecasting; Long Short-Term Memory (LSTM); Bidirectional LSTM; Time-Series Prediction; Precision Agriculture; Edge Computing.

1. Introduction

Greenhouse cultivation has emerged as a sustainable alternative to open-field agriculture, offering protection against climatic variability and enabling year-round production [1] [2]. Within these controlled environments, microclimatic variables such as temperature, relative humidity, and CO_2 concentration interact nonlinearly, collectively influencing plant growth, yield, and quality [3][4]. IoT-enabled vertical farming systems have demonstrated improved resource efficiency and yield prediction accuracy when combined with regression-based

machine learning models, particularly in hydroponic and aeroponic environments [5]. Similarly, hybrid modeling approaches that integrate physics-based formulations with machine learning algorithms have shown strong performance in estimating crop growth dynamics and resource consumption in aeroponic lettuce systems, highlighting the importance of coupling mechanistic understanding with data-driven learning [6]. In controlled-environment agriculture, deep learning frameworks have become central to optimizing production systems by enabling predictive control and adaptive environmental regulation . [7]. Accurate prediction and regulation of these

*Corresponding Author

Siman Emmanuel*

Email: siman.emma@fuwukari.edu.ng

This is an open access article under the [CC BY-NC](#) license



variables remain critical yet challenging due to their interdependence and dynamic fluctuations [8] [9]. Early methods relied on statistical modeling and sensor-based monitoring but were limited in handling missing data and capturing temporal dependencies [2][10]. The integration of IoT technologies with microcontrollers like ESP8266/ESP32 and NodeMCU enabled real-time sensing, automated irrigation, and climate control, improving resource efficiency and crop performance. However, early deployments often faced short evaluation periods, scalability issues, and reliance on stable connectivity, limiting generalization to different crops and operational scales [11]; [12]; [13]. Deep learning has significantly advanced greenhouse microclimate prediction. Architectures evolved from Feedforward Neural Networks (FNNs) to Recurrent Neural Networks (RNNs), and then to Long Short-Term Memory (LSTM) networks, which address the vanishing gradient problem and enable long-term temporal dependency modeling [14] [15]. Hybrid LSTM/Bi-LSTM frameworks leverage bidirectional memory for enhanced prediction accuracy, capturing complex temporal and multivariate patterns in greenhouse data [6]; [8]. Combining AI, IoT, and energy management frameworks further optimizes greenhouse operations, reducing electricity consumption, grid dependency, and resource wastage while maintaining microclimate stability [16]; [17]; [18]. Ensemble methods, including Random Forest and Gradient Boosting, alongside ANN and Bi-LSTM networks, enhance real-time prediction and decision-making for adaptive climate control [8]. Hybrid GA-optimized LSTM with fuzzy logic controllers improves actuator response, ensuring energy-efficient regulation of temperature and humidity under dynamic greenhouse conditions [19].

Despite these advances, challenges persist in standardizing datasets, achieving robust generalization across regions and crops, and deploying scalable, real-time AI-enabled systems [10] [11] [20]. Precision agriculture further emphasizes the integration of remote sensing and artificial intelligence for environmental monitoring and decision support, demonstrating that deep learning models significantly outperform traditional statistical approaches in complex agricultural environments [21]. In addition, edge and fog computing architectures have been proposed to support real-time agricultural analytics, reducing latency and improving scalability in IoT-enabled systems [22]. Time-series forecasting models such as LSTM networks have also been widely adopted due to their ability to capture long-term dependencies in nonlinear environmental data [23]. Emerging strategies, including edge computing, digital twins, and federated learning, offer potential for low-latency, adaptive decision-making but remain underexplored in precision greenhouse applications [24] [25]. Energy efficiency and sustainability considerations are increasingly integrated into smart agricultural systems, particularly through optimized thermal and solar energy utilization in

controlled environments [26]. However, despite these advancements, challenges remain in achieving robust, scalable, and generalizable predictive models capable of handling multivariate greenhouse dynamics under real-world variability. Addressing these gaps is essential to realize hybrid LSTM/Bi-LSTM frameworks capable of optimizing greenhouse microclimates, enhancing crop productivity, resource efficiency, and sustainability [27] [28]. This study presents a real-time IoT-enabled greenhouse monitoring and forecasting framework integrated with hybrid deep learning architectures for multivariate microclimate prediction. The main contributions are as follows:

- i. First, a unified IoT–edge–cloud architecture is developed for continuous acquisition, preprocessing, and transmission of high-resolution greenhouse environmental data. The framework supports low-latency monitoring of key microclimate variables, enabling structured temporal datasets suitable for deep learning-based forecasting.
- ii. Second, a hybrid sequence learning strategy combining Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) networks is introduced for multivariate time-series prediction of greenhouse temperature, humidity, soil moisture, and light intensity. The bidirectional learning mechanism improves temporal dependency extraction, particularly under nonlinear and rapidly fluctuating environmental conditions.
- iii. Third, a systematic comparative evaluation is conducted against baseline deep learning configurations, demonstrating the relative advantages of bidirectional temporal encoding in reducing prediction error and improving coefficient of determination under normalized sensing conditions.
- iv. Fourth, an end-to-end predictive workflow is established, linking sensor-level data acquisition with model inference for real-time microclimate estimation, providing a computational basis for adaptive greenhouse management and decision support.
- v. Finally, the research demonstrates the feasibility of deploying lightweight deep learning models within IoT-assisted agricultural environments, highlighting the trade-off between predictive accuracy and computational overhead, which is critical for edge-based deployment scenarios.

Section 2 highlights the challenges in predicting greenhouse microclimate conditions, emphasizing the need for deep learning models such as LSTM and Bi-LSTM to overcome vanishing gradient issues and capture long-term dependencies in time-series data. Section 3 details the methodology at the Faculty of Agriculture, Federal

University Wukari, where IoT sensors continuously monitor temperature, light status, soil moisture, and humidity. Collected data are processed using LSTM and Bi-LSTM models, with hyperparameter tuning and performance evaluated via R^2 , MAE, RMSE, and MSE. Section 4 describes the real-time data acquisition, preprocessing on edge devices, and cloud-based storage for further analysis. Section 5 presents the model development, including LSTM and Bi-LSTM architectures with forget, input, and output gates, and a hybrid model capturing both spatial and temporal dependencies. Section 6 reports the performance evaluation, demonstrating that Bi-LSTM outperforms LSTM in predictive accuracy for greenhouse microclimate forecasting.

2. Background of study

The evolution of greenhouse environmental modeling has progressed from statistical and empirical approaches to advanced machine learning and deep learning frameworks. Early developments in environmental prediction systems demonstrated the use of machine learning for dynamic system modeling, including methane emission prediction and environmental process forecasting, establishing the foundation for AI-driven environmental analysis [29]. Initial attempts using Recurrent Neural Networks (RNNs) effectively predicted greenhouse air temperature (RMSE 1.50 °C; MAPE 4.91%; R^2 0.965), but were constrained by limited data handling and small-scale deployment [2][4]. Advancements in smart-greenhouse platforms integrated IoT sensors, renewable power, and internet connectivity, enabling real-time environmental monitoring and predictive control. However, these systems suffered from seasonal drift, high maintenance demands, and limited applicability across diverse crops and larger operational scales [10]; [11]. Early LSTM/Bi-LSTM forecasting demonstrated improved predictive accuracy, yet short evaluation periods and reliance on stable connectivity limited generalizability [11]. Recent research has also demonstrated the effectiveness of LSTM-based models in evapotranspiration and humidity prediction for greenhouse systems, highlighting their suitability for moisture regulation and irrigation control [30]. Comparative studies further confirm that ensemble and hybrid deep learning models consistently outperform classical ANN and regression-based approaches in modeling greenhouse climate variables [31]. Subsequent research leveraged low-cost IoT-based drip irrigation systems and machine learning to optimize water use, energy consumption, and crop yield [12]. NodeMCU and ESP8266/ESP32 microcontrollers enabled automated irrigation, temperature, and air circulation, achieving up to 30% water savings and 15–18% yield improvement. Machine learning models including Random Forest (RF), Artificial Neural Networks (ANN), and XGBoost provided prediction accuracies up to 99.96%, though adoption remained constrained by high costs, connectivity challenges,

cybersecurity risks, and the absence of standardized benchmarks [12]; [17]; [18]; [13]. Hybrid GA-optimized LSTM coupled with fuzzy logic controllers further enhanced precision in temperature and humidity regulation, reducing energy waste while responding to dynamic microclimatic variations [19]. Parallel efforts addressed energy management and sustainability in controlled environments. IoT-driven demand-side management frameworks combining XGBoost and LSTM achieved reduced grid dependency (7.1%) and electricity cost savings (17.3%), demonstrating scalable and economically viable energy optimization while maintaining microclimatic integrity [16]. In controlled agricultural systems, deep learning techniques have been extensively applied to microclimate prediction and crop monitoring, showing strong performance improvements in temperature and humidity forecasting compared to conventional models [32]. Hybrid architectures combining recurrent neural networks and LSTM models have further enhanced predictive accuracy in greenhouse environments by capturing both short-term fluctuations and long-term dependencies [33]. Machine learning models (Bi-LSTM, ANN, GBM, RF) improved predictive capabilities for temperature, humidity, and CO₂, with ensemble and deep learning methods outperforming conventional models, highlighting the importance of sensor-specific features in precision agriculture [8]. Broader applications integrated AI and IoT across precision agriculture, eco-city initiatives, and livestock management. AGI frameworks for IoT-enabled systems demonstrated transformative potential in intelligent automation and decision-making, though limitations in computing resources, communication complexity, and security persisted [34] [35]. Smart precision irrigation and aeroponic cultivation employed hybrid ML-physics-based models to predict fresh weight, leaf area, nitrate levels, and water use, achieving robust performance for growth-related metrics while highlighting constraints in nitrate and water prediction due to limited training data [6][28]. Eco-city frameworks utilized UAVs, GIS, and automated irrigation to optimize urban food production while promoting resource efficiency, environmental sustainability, and alignment with SDGs 2, 11, 12, and 13 [25]. Integration of AI-IoT in precision livestock farming enhanced indoor air quality monitoring and emission control. AIoT systems outperformed conventional methods but faced challenges in sensor durability, intelligent control design, cost, and energy efficiency, necessitating future work in low-cost sensors, multi-variable control strategies, and end-edge-cloud AIoT integration [36]. Sensorics combined with IoT also improved environmental monitoring for water quality, biodiversity, habitat, and climate assessments, though challenges remained in sensor durability, data standardization, and large-scale deployment [37]. Further, real-time data from sensors, drones, and satellites, combined with AI analytics, enabled adaptive irrigation, pest control, and ecosystem monitoring, while predictive modeling strengthened financial transparency, risk assessment, and

subsidy optimization in agriculture [38]; [39]. Systematic reviews emphasized that smart irrigation, deep learning-based forecasting, and IoT integration significantly improve water use efficiency, crop yield, and resource management. Yet, fragmented approaches, inconsistent sustainability metrics, and limited closed-loop edge implementations present gaps for future research in scalable, benchmarked, and sustainable smart agriculture[40] [20][24]. Abiotic stress management in wheat further illustrates the convergence of genomics, physiology, high-throughput phenotyping, and precision agronomic practices, enabling resilience under drought, salinity, temperature extremes, and nutrient deficiencies while maintaining productivity and environmental sustainability [27]. Beyond agricultural

systems, stochastic environmental modeling and hydrological cycle analysis have provided additional insights into uncertainty handling in complex systems, reinforcing the need for probabilistic and data-driven approaches in environmental prediction tasks [41]. These developments collectively establish a strong foundation for integrating IoT-based sensing with advanced deep learning architectures for precision greenhouse management. Collectively, this research reflect a progressive evolution from early sensor-based predictions to integrated AI-IoT frameworks, addressing prior limitations in scalability, generalization, and operational efficiency, and charting the trajectory for next-generation precision agriculture and controlled-environment systems.

Table 1: Integrated Synthesis of Greenhouse Microclimate Prediction and Smart Agriculture Systems

Category	Reference	System Type	Objective	Model Architecture /	Key Inputs	Performance Summary	Core Limitation
Classical Time-Series Models	[2]	Greenhouse climate system	Temperature prediction	RNN	Temperature, humidity	$R^2 \approx 0.965$	Weak scalability, short temporal memory
	[23].	Environmental forecasting	Time-series modeling	RNN variants	Multivariate signals	High predictive ability	Limited nonlinear generalization
	[42]	Greenhouse irrigation system	Evapotranspiration prediction	LSTM	Moisture variables	High accuracy	Limited external validation
Deep Learning Climate Models	[32].	Agricultural microclimate system	Temperature/humidity forecasting	Deep learning models	Sensor-based climate data	Outperforms classical ML	High computational cost
	[33]	Greenhouse system	Climate prediction	RNN–LSTM hybrid	Environmental time-series	Improved accuracy vs ANN	Limited deployment scope
	[8]	Precision agriculture	Multivariate prediction	Bi-LSTM, ANN, RF	Temp, humidity, CO ₂	Bi-LSTM best performance	No real-time integration
Hybrid & Optimized Models	[6]	Aeroponic system	Growth prediction	Physics-informed ML	Environmental + physiological data	High robustness	Weak nitrate/water prediction
	[19]	Greenhouse control	Climate regulation	GA-optimized LSTM + fuzzy logic	Temperature, humidity	Improved actuator response	High computational complexity
	[16]	Energy-efficient greenhouse	Energy optimization	XGBoost + LSTM	Load + climate variables	17.3% cost reduction	Not crop-centric
IoT & Edge-Based Systems	[17]	Smart greenhouse	Environmental control	IoT + DL system	Sensor network data	Improved efficiency	Weak predictive depth
	[5]	Vertical farming	Yield prediction	IoT + regression ML	Environmental variables	Improved yield estimation	Weak temporal modeling
	[12]	Irrigation	Water optimization	IoT	Soil moisture,	30% water	Poor

		system		(ESP32/ESP8266)	temp	savings	generalization
	[11]	Greenhouse monitoring	Climate sensing	IoT sensor network	Temp, humidity	Real-time monitoring	Connectivity instability
	[22]	Edge agriculture	Real-time analytics	Edge-cloud architecture	Sensor fusion data	Reduced latency	Limited large-scale validation
	[36]	AIoT agriculture	Environmental monitoring	AIoT system	Air quality, emissions	Improved monitoring	Sensor durability issues
Proposed Framework	This Research	Real-time greenhouse system	Microclimate forecasting	IoT + LSTM + Bi-LSTM ensemble	Temp, humidity, soil moisture, light	R² ≈ 0.9981, RMSE ≈ 0.0110	Needs long-term multi-location validation

The integrated Table 1 presents a unified view of the evolution of greenhouse microclimate prediction and smart agriculture systems, spanning classical statistical models, deep learning approaches, hybrid/optimized frameworks, and IoT-edge-based architectures. Early studies based on RNN and basic LSTM models demonstrate the ability to capture temporal dependencies in environmental variables such as temperature and humidity, but they are generally limited by short-term datasets, weak generalization, and poor scalability. Subsequent deep learning and hybrid models, including Bi-LSTM, RNN-LSTM combinations, and physics-informed machine learning, significantly improve predictive accuracy by better capturing nonlinear and multivariate interactions, although they often remain computationally intensive and lack full real-time deployment capability. In parallel, IoT and edge computing systems enable real-time environmental sensing and control, improving operational efficiency in greenhouse management, but these systems typically rely on shallow predictive models and do not fully exploit advanced temporal learning. Therefore, the table highlights a clear progression from isolated predictive models to integrated IoT-AI frameworks, while revealing a persistent gap in fully unified, real-time, and scalable greenhouse forecasting systems, which is directly addressed by the proposed IoT-LSTM-BiLSTM ensemble approach in this study.

3. Methodology

The proposed framework integrates IoT-based sensing with hybrid LSTM and Bi-LSTM architectures to model greenhouse microclimate dynamics. The adoption of deep learning is motivated by its proven ability to handle nonlinear temporal dependencies in environmental datasets, particularly in controlled agriculture systems where multivariate interactions dominate system behavior [23]. The hybrid modeling strategy is further supported by studies demonstrating that combined physics-informed and machine learning approaches improve predictive robustness in agricultural systems, particularly in aeroponic and hydroponic crop production environments [6]. Similarly, IoT-integrated regression models have shown strong performance in optimizing crop yield prediction in vertical

farming systems, reinforcing the relevance of data-driven modeling in controlled environments [5]. Feature selection and preprocessing strategies adopted in this research align with established practices in precision agriculture, where remote sensing and sensor fusion techniques are used to improve data quality and predictive stability [21]. Additionally, edge computing architectures have been shown to enhance real-time responsiveness in IoT-based agricultural systems, justifying their integration into the proposed framework [22]. The use of LSTM and Bi-LSTM architectures is further supported by their proven effectiveness in time-series forecasting applications across environmental systems, where they outperform traditional models in capturing sequential dependencies and nonlinear relationships [23].

3.1 Methodology Overview

The proposed framework integrates IoT-based sensing, edge preprocessing, and ensemble deep learning for greenhouse microclimate prediction. Environmental variables (temperature, humidity, soil moisture, and light intensity) were acquired using IoT sensors and transmitted to an edge device for real-time preprocessing. Missing values were handled using mean imputation, while outliers were removed using Z-score analysis with a threshold of $|z| > 3$. All features were normalized using Min-Max scaling to ensure uniform input distribution. Time-series data were transformed into supervised learning sequences using a sliding window approach. Two deep learning models, LSTM and Bi-LSTM, were independently trained to capture forward temporal dependencies and bidirectional contextual information, respectively. To improve robustness, a weighted ensemble fusion strategy was applied:

$$\hat{y} = w_1 \hat{y}_{LSTM} + w_2 \hat{y}_{BiLSTM}$$

where weights were optimized based on validation performance.

Model training was performed using the Adam optimizer with MSE loss, a learning rate of 0.0001, batch size of 5, and 20 epochs. Performance was evaluated using R², MAE, RMSE, and MSE. Statistical validation included paired t-test, Wilcoxon signed-rank test, 95% confidence intervals,

and multi-run stability analysis ($k \geq 10$), ensuring robustness and reproducibility of results.

3.2 Research Area/Site

The research will be conducted at the Faculty of Agriculture, Federal University Wukari, located in a tropical savanna climate region. The university's greenhouse facilities, designed for experimental farming, are equipped with modern crop management systems. These facilities are ideally suited for testing IoT sensors and deep learning models. The controlled environment of the greenhouses, with variability in Temperature, Light status, Soil Moisture and Humidity, offers a dynamic setting for analyzing greenhouse microclimates. This setting provides real-world data essential for the development and evaluation of predictive models aimed at optimizing greenhouse climate conditions for improved crop production.

3.2.1 Plant Cultivation

The greenhouse experiments involved cultivating cucumber, tomato, and pepper, chosen for their economic value and sensitivity to microclimatic changes. IoT sensors were strategically positioned around these plants to capture environmental parameters such as temperature, humidity, light intensity, and soil moisture. The collected data allowed the LSTM and Bi-LSTM models to learn crop-specific responses, enabling precise optimization of greenhouse conditions for each crop type and improving overall growth and yield.

The aim of this research is to create a computational framework for predicting greenhouse microclimate using LSTM and Bi-LSTM models. The proposed methodology involves selecting relevant characteristics from environmental parameters and standardizing the data. The model understands temporal interdependencies through input sequences and fine-tuning hyperparameters. IoT sensors (temperature, humidity, soil moisture, and light intensity) were calibrated for accurate data collection. The data was processed using a noise-filtering algorithm and standardized with min-max normalization to ensure consistency. The LSTM and Bi-LSTM models were

optimized using Bayesian optimization for hyperparameter tuning. This process enabled the fine-tuning of key parameters such as the learning rate and batch size, allowing for better model performance and reduced overfitting.

3.2.2 Materials and Methods

The materials and methods for greenhouse cultivation involved establishing the greenhouse on 21 August 2025, with bed preparation completed on 25 August 2025 using a soil–poultry manure mix, watered for two weeks prior to planting. Four crops were cultivated in designated blocks: Tomato (Block A), Cucumber (Block B), Pepper (Block C), and Cabbage (Block D), spaced at 60 × 50 cm per plant. Planting commenced on 22 October 2025, with nursery beds for Pepper, Cabbage, and Tomato established on 4 November 2025, followed by transplanting of Pepper and Tomato on 15 November 2025. Growth was monitored, noting cucumber germination on 25 October 2025 and flowering on 15 December 2025, while Pepper and Tomato flowered on 16 December 2025. Harvesting of cucumber occurred on 7 January 2026, yielding two fruits with recorded weights, lengths, and girths of 0.005 kg (25 cm, 20 cm) and 0.002 kg (20 cm, 17 cm), respectively, providing quantitative data for performance assessment and analysis

Table 2 the greenhouse crop production report summarizes all management activities, key dates, and harvest outcomes for Tomato, Cucumber, Pepper, and Cabbage cultivated under controlled conditions. The greenhouse was established on 21 August 2025, with beds prepared and enriched with poultry manure by 25 August 2025, and crops planted on 22 October 2025. Cucumber germinated on 25 October 2025, with nursery beds for Pepper, Tomato, and Cabbage established on 4 November 2025, followed by transplanting of Pepper and Tomato on 15 November 2025. Flowering occurred in mid-December, and cucumber was harvested on 7 January 2026, yielding two marketable fruits weighing 0.005 kg and 0.002 kg, with lengths of 25 cm and 20 cm and girths of 20 cm and 17 cm, respectively. These records demonstrate effective crop establishment and provide a reliable baseline for analysis, predictive modeling, and performance assessment of greenhouse operations

Table 2: Greenhouse crop production report

I Greenhouse Setup and Preparation		V Harvesting	
Activity	Date/Description	Crop	Date of Harvesting
Mounting of greenhouse	21/08/2025	Cucumber	07/01/2026
Bed preparation	25/08/2025		
Soil preparation	Soil mixed with poultry manure and water for 2 weeks before planting		
II Crops and Field Layout		VI Fruits and Yield Measurements (Cucumber)	
		Number of fruits arrested: 2	
		Measurements and weights are presented below.	

Block	Crop	Spacing
A	Tomato	60 x 50 cm (each)
B	Cucumber	60 x 50 cm (each)
C	Pepper	60 x 50 cm (each)
D	Cabbage	60 x 50 cm (each)

Fruit No.	weight (kg)	weight (g)	length (cm)	Girth (cm)
1 st	0.005	5	25	20
2 nd	0.002	2	20	17

III Planting and Germination

Activity	Date/Details
Date of planting (all crops)	22/10/2025
Date of germination (Cucumber)	25/10/2025
Flowering of Cucumber	15/12/2025

IV Nursery and Transplanting

Activity	Date/Details
Nursery bed (Pepper, Cabbage, Tomato)	4/11/2025
Transplanting Pepper and Tomato	15/11/2025
Flowering of Pepper and Tomato	16/12/2025

VII Summary of Key Dates

Activity	Date
Mounting of greenhouse	21/08/2025
Bed preparation	25/08/2025
Soil mixed with poultry manure (2 weeks before planting)	25/08/2025
Date of planting (all crops)	22/10/2025
Date of germination (Cucumber)	25/10/2025
Nursery bed (Pepper, Cabbage, Tomato)	4/11/2025
Transplanting Pepper and Tomato	15/11/2025
Flowering of Cucumber	15/12/2025
Flowering of Pepper and Tomato	16/12/2025
Harvesting of Cucumber	07/01/2026

3.3. Data Collection:

Data collection will involve the deployment of IoT sensors within the greenhouse facilities to monitor key environmental parameters, including Temperature, Light status, Soil Moisture and Humidity levels at intervals of less than one minute. The collected data will be preprocessed on edge devices, such as Raspberry Pi and Jetson, to filter out noise and handle missing values. After preprocessing, the data will be transmitted to cloud servers, specifically ThingSpeak, for centralized storage and further analysis. This real-time data collection and processing framework will enable the accurate monitoring of greenhouse microclimate conditions, essential for evaluating and optimizing environmental control strategies. Sensor calibration was conducted prior to data collection using standard industry methods, ensuring high accuracy of measurements. Missing values in the dataset were imputed using mean imputation, and outliers were removed using a Z-score threshold of 3, ensuring the integrity of the data used for training the models.

4 Model Development

LSTM (Long Short-Term Memory) and Bi-LSTM (Bidirectional LSTM) models will be developed to forecast greenhouse climate conditions, with a focus on essential parameters like as temperature, light status, soil moisture, and humidity. In addition, a hybrid LSTM and Bi-LSTM architecture will be tested to capture spatial and temporal dependencies in the data. This hybrid technique aims to increase the model's ability to handle complex, multidimensional datasets. Bayesian optimization will be utilized to determine the optimal model parameter choices, resulting in the highest predictive accuracy and generalization performance. These models are created and

validated to ensure their efficacy in real-time climate forecast in a greenhouse environment.

4.1 The Long Short-Term Memory (LSTM)

The vanishing gradient problem in RNNs is successfully overcome using LSTM models. The use of LSTM models preserves long-term input dependence while increasing memory capacity. LSTM memory selectively retains or ignores information. An LSTM model captures important input properties and maintains this information for an extended period of time. The allocated weight values throughout the training procedure influence the decision to maintain or reject information (Imrana et al., 2021). As a result, an LSTM model identifies key traits and utilizes weight values to decide whether to keep or discard information. LSTM models consist of three gates: the forget gate, the input gate, and the output gate. The forget gate assesses the importance of data. The input gate decides whether or not to add fresh data. The value to be kept in memory is decided by the output gate. Whether a value needs to be updated is decided by the forget gate. Potential new values are generated by the input gate. The new candidate value is kept in memory while the old value is forgotten.

- i. The Forget Gate: The output of the forget gate determines whether to keep or remove the current value. The decision is mostly determined by the values of h_{t-1} and X_t . The forget gate's final output, denoted by the symbol ft , ranges from 0 to 1, with 0 indicating complete elimination of the learned value and 1 indicating preservation of the entire value. The following equation represents the calculation of this output:

$$ft = \sigma(w_{if}Xt + b_{if} + w_{hf} h_{t-1} + b_{hf}) \quad (1)$$

where W is weight matrices b_f represents a constant known as the bias value x_t is the input at time t , σ is the sigmoid activation function,

- ii. In contrast, the input gate decides whether or not to add fresh data to the LSTM memory. A sigmoid layer and a "tanh" layer make up this gate's two layers. The "tanh" layer creates a vector of possible new values to be added to the LSTM memory, while the sigmoid layer determines which values need to be updated. The following computations establish the outputs of these two layers:

$$it = \sigma(W_{ii}Xt + b_{ii} + W_{hi} h_{t-1} + b_{hi}) \quad (2)$$

In this scenario, ct represents a vector of candidate values that could be added to the LSTM memory, as well as the decision on whether or not to update a value. These two layers collaborate to make it easier to update LSTM memory. To forget the current value, the forget gate layer multiplies the previous value ($ct - 1$) and adds the new candidate value ($c't$). This procedure can be quantitatively described by the following equation:

$$ct = ft \odot c_{t-1} + i_t \odot g_t \quad (3)$$

In this equation, $\odot$ denotes element-wise multiplication, ft represents the output of the forget gate, which falls within the range of 0 to 1. A value of 0 signifies complete elimination of the current value, while a value of 1 implies complete preservation of the value.

- iii. The output gate.

To determine which portion of the LSTM memory contributes to the output, this gate initially employs a sigmoid layer. The numbers between -1 and 1 are then

mapped using a non-linear tanh function. Lastly, the output of a sigmoid layer is multiplied by the result. The formulas to calculate the output are represented by the following equation.

$$ot = \sigma(w_{io}Xt + b_{io} + w_{ho} h_{t-1} + b_{ho}) \quad (4)$$

$$ht = o_t \odot \tanh(c_t) \quad (5)$$

where o_t is the output value, and ht is its representation as a value between -1 and 1 .

4.2. The Bidirectional Long Short-Term Memory (BLSTM)

Bidirectional Long Short-Term Memory (BLSTM) is an extension of the classic Long Short-Term Memory (LSTM) framework that processes incoming data both forwards and backwards. The hybrid of bidirectional LSTM is known as BLSTM. Unlike RNN, which only employs past contexts, BLSTM may access both past and future contexts given an input sequence. BLSTM consists of two hidden layers that are concatenated together. The first layer processes the input data forward, while the second layer processes it backward. By employing the LSTM twice, long-term dependencies are better learned, increasing the model's accuracy.

4.3. Concatenation of Forward and Backward Outputs

$$htBLSTM = (ht + h't) \quad (6)$$

Here,

$ht - 1$ and $h't + 1$ are the hidden state from earlier time steps; i_t, f_t are the input, forget, cell, and output gates for the forward LSTM, i'_t, f'_t, g'_t are the input, forget, cell, and output gates for the reverse LSTM; ct and $c't$ are the cell states of forward and backward LSTMs; and ht and $h't$ are the hidden states of forward and backward LSTMs. As a result, an LSTM and Bi-LSTM model were utilized to investigate greenhouse microclimate.

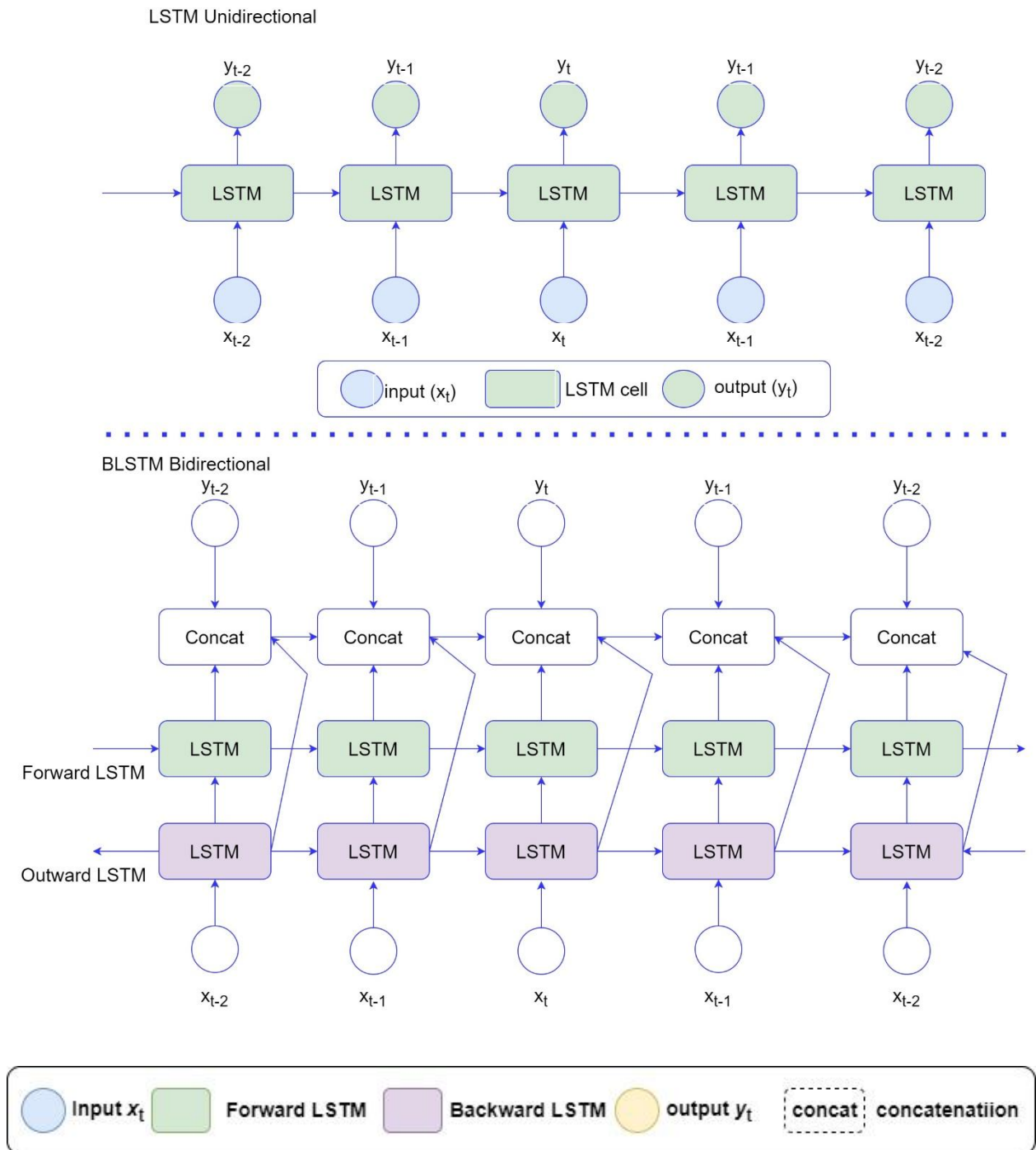


Figure 1 (a) Schematic diagram of the LSTM and BLSTM Architecture

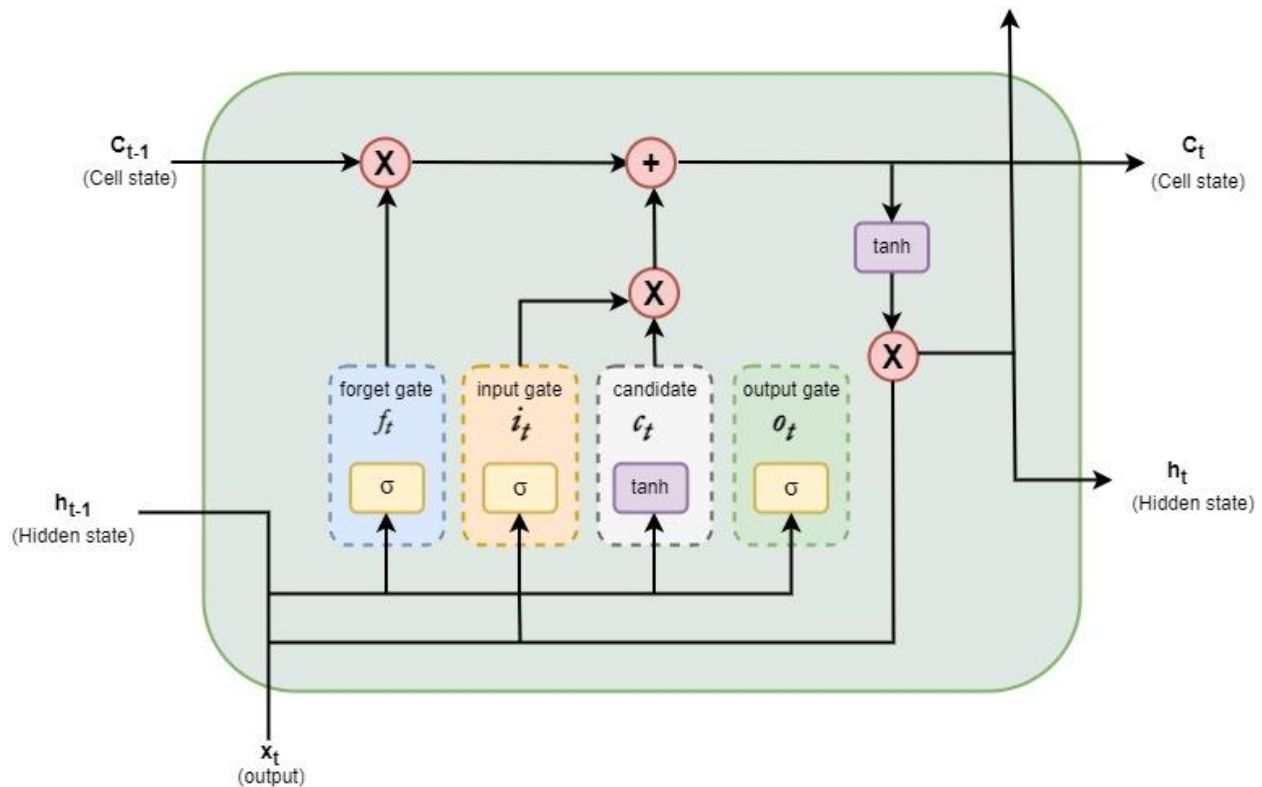


Figure 1 (b) Long Short-Term Memory (LSTM) cell architecture diagram

Figure 1(a) illustrates the internal structure of an LSTM unit used in recurrent neural networks for sequential data processing. The model takes the previous hidden state (h_{t-1}), previous cell state (c_{t-1}), and current input (x_t) to regulate information flow through three main gates: the forget gate, input gate, and output gate. The forget gate determines which parts of the previous memory should be discarded, while the input gate controls how much new information is stored in the cell state through a candidate update. The cell state acts as the long-term memory carrier, being updated by combining retained past information and newly added information. Finally, the output gate determines the portion of the internal state that is exposed as the hidden state (h_t), which is passed to the next time step or output layer. This gating mechanism enables the network to effectively capture long-range dependencies in sequential data such as time series, biosignals, or environmental datasets. The Figure (b) represents a Long Short-Term

Memory (LSTM) and Bidirectional LSTM (BiLSTM) neural network architecture used for sequential data modelling. It shows how input data passes through memory-based recurrent units that are designed to retain important information over time while filtering irrelevant signals using internal gating mechanisms. Each LSTM unit regulates information flow through input, forget, and output gates, enabling the network to capture long-term dependencies in time-dependent datasets. In the bidirectional configuration, the sequence is processed in both forward and backward directions, allowing the model to integrate past and future contextual information simultaneously before producing a final output. This structure is particularly effective for regression or classification tasks involving temporal or ordered data, as it improves feature representation and prediction accuracy compared to standard feedforward neural networks.

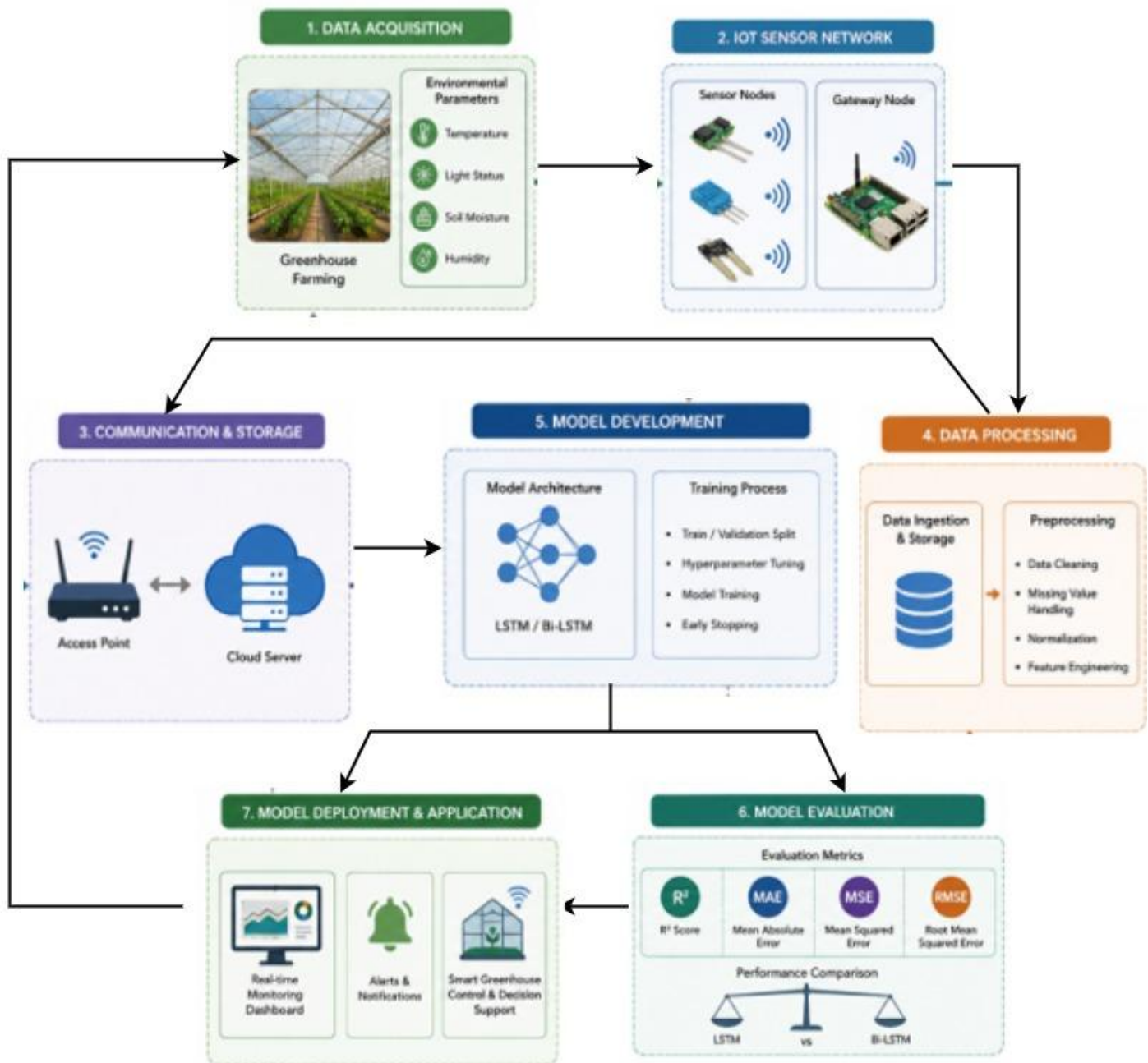


Figure 2: Framework for the model development

The four basic nodes that comprise the Internet of Things architecture are sensor nodes, gateway nodes, processor and storage nodes, as illustrated in Figure 2. Sensor nodes receive data from each sensor and regularly relay it to the gateway node, either directly or through other sensor nodes, depending on their distance from it. They built up a wifi network in the greenhouses so that they could communicate with one another. In addition to the sensor nodes, a gateway node is deployed to permit communication between the

greenhouse and the Internet (cloud server). The gateway node supports two communication interfaces: Wi-Fi to stay in touch with the sensor nodes and a cellular interface to provide information to the processing and storage node. In order to understand the conditions of the greenhouse and operate the various sensor and actuator devices, a farmer has remote access to the data collection. To send all collected data to the cloud server, the gateway node has a cellular network interface.

Algorithm 1: Ensemble-Based LSTM–Bi-LSTM Framework for Real-Time IoT-Enabled Greenhouse Microclimate Forecasting

Require: Input sequence X_t (Temperature, Light, Soil Moisture, Humidity), weights W , biases b , previous hidden states $h_{t-1}^{LSTM}, h_{t-1}^{Bi-LSTM}, h_{t+1}^{Bi-LSTM}$, previous cell states $C_{t-1}^{LSTM}, C_{t-1}^{Bi-LSTM}, C_{t+1}^{Bi-LSTM}$, learning rate, epochs, batch size

Ensure: Predicted greenhouse microclimate $\hat{Y}_t$

```

1: Initialize LSTM and Bi-LSTM parameters (weights, biases, gates)
2: Normalize input data  $X_t$  using Min-Max normalization
3: Divide dataset into training and testing sets (80% train, 20% test)
4: for each epoch do
5:   for each batch in training data do
6:     LSTM forward pass:
7:       for each time step  $t$  do
8:         Compute forget, input, candidate cell, and output gates (as in Algorithm 1)
9:         Update LSTM hidden and cell states
10:      end for
11:    Bi-LSTM forward-backward pass:
12:      for each time step  $t$  do
13:        Compute forward and backward gates (as in Algorithm 2)
14:        Update forward and backward hidden and cell states
15:        Concatenate forward and backward hidden states:  $h_t^{Bi} = [h_t^f; h_t^b]$ 
16:      end for
17:    Combine LSTM and Bi-LSTM outputs:  $h_t^{Hybrid} = h_t^{LSTM} + h_t^{Bi}$  or weighted sum
18:    Compute predicted output  $\hat{Y}_t$  from  $h_t^{Hybrid}$ 
19:    Compute loss between  $\hat{Y}_t$  and actual  $Y_t$ 
20:    Backpropagate loss through both LSTM and Bi-LSTM components using Adam optimizer
21:  end for
22: end for
23: Apply Dropout regularization (0.3) to all layers
24: Apply ReLU activation in hidden layers and SoftMax in output layer
25: Evaluate model performance using  $R^2$ , MAE, MSE, RMSE
26: Output predicted greenhouse microclimate  $\hat{Y}_t$ 

```

Algorithm 1 formalizes an ensemble-based temporal learning strategy integrating independent LSTM and Bi-LSTM learners with weighted prediction fusion for real-time greenhouse microclimate forecasting. The hybrid LSTM–Bi-LSTM framework was selected due to its ability to capture complementary temporal representations, where the LSTM model effectively learns forward temporal dependencies in environmental variables, while the Bi-LSTM model enhances feature representation by incorporating both forward and backward contextual information within the input sequence. This dual-model configuration improves sensitivity to abrupt and nonlinear fluctuations in temperature, light intensity, soil moisture, and humidity, which are common in dynamic greenhouse environments. The weighted ensemble fusion further enhances predictive stability by combining the strengths of both models, resulting in robust forecasting performance with high accuracy ($R^2 \approx 0.9981$) and low error margins (MAE and RMSE), outperforming the individual LSTM and Bi-LSTM models. Moreover, the proposed ensemble framework aligns seamlessly with the IoT-based data

acquisition architecture, enabling real-time forecasting and adaptive greenhouse climate regulation. This integration supports improved crop growth conditions and optimized resource utilization. Consequently, Algorithm 1 provides a balanced and scalable solution for capturing complex multivariate temporal interactions and achieving stable, high-fidelity microclimate predictions.

5 Dataset

This section describes the datasets used for model development, validation, and deployment, including both the benchmark greenhouse dataset and the original IoT-enabled greenhouse dataset generated in this research. Despite the relatively short acquisition period, the dense temporal resolution (10-minute sampling) provides sufficient information richness for training deep sequential models, which are known to perform effectively under high-frequency time-series conditions even with moderate duration datasets.

5.1 Dataset Development and Validation Strategy

A multi-stage dataset strategy was adopted to ensure robust model development, validation, and real-world deployment. In the first stage, a publicly available greenhouse dataset reported by [3] was employed for preliminary model development, benchmarking, feature engineering, and hyperparameter optimization. This benchmark dataset provided a standardized platform for evaluating the predictive capability of the proposed LSTM and Bi-LSTM architectures and for identifying optimal model configurations. In the second stage, an original greenhouse dataset was generated through the implementation of an IoT-enabled greenhouse monitoring system at the Faculty of Agriculture, Federal University Wukari, Nigeria. Environmental parameters, including temperature, humidity,

soil moisture, and light intensity, were continuously acquired using integrated sensor nodes during the cultivation of cucumber, tomato, and pepper crops. The collected data were transmitted through the proposed IoT-edge-cloud framework and stored for subsequent analysis and model validation. In the final stage, the enhanced model developed from the benchmark dataset was validated and deployed using the locally acquired greenhouse dataset. This enabled the assessment of model generalization, forecasting accuracy, and practical applicability under real greenhouse operating conditions. The adopted strategy combines the advantages of benchmark-based model optimization with real-world validation, thereby improving the reliability and scalability of the proposed intelligent greenhouse monitoring framework as in Table 3.

Table 3 summarizes the dataset development and validation workflow adopted in this research

Phase	Dataset Source	Purpose	Outcome
Phase I	[3] Dataset	Model development, benchmarking, feature engineering, and hyperparameter optimization	Enhanced LSTM/Bi-LSTM framework developed
Phase II	Federal University Wukari Greenhouse Dataset	Validation using locally acquired IoT sensor data from cucumber, tomato, and pepper cultivation	Model robustness and generalization verified
Phase III	Real-Time Greenhouse Deployment	Online forecasting of temperature, humidity, soil moisture, and light intensity	Demonstrated applicability for precision greenhouse management

5.2 Feature Engineering

Feature engineering was performed to enhance the predictive capability of the proposed deep learning models by transforming raw greenhouse sensor measurements into structured inputs suitable for time-series forecasting. Environmental parameters acquired from the IoT-enabled greenhouse system, including temperature, humidity, soil moisture, and light intensity, were first subjected to data preprocessing to improve data quality and consistency.

Missing observations resulting from sensor communication interruptions were handled through mean-value imputation, while abnormal measurements were identified and removed using Z-score analysis with a threshold of ± 3 . To eliminate scale differences among variables and facilitate efficient model convergence, all features were normalized using Min-Max scaling, which transformed the data into the range of 0 to 1.

Following preprocessing, the dataset was divided into training and testing subsets using an 80:20 ratio. Time-series sequences were then generated using a sliding-window approach to preserve temporal dependencies among successive observations. The engineered feature set consisted of temperature, humidity, soil moisture, and light intensity measurements collected from the greenhouse environment, which served as input variables for the LSTM and Bi-LSTM models. These features were selected because

they directly influence greenhouse microclimate dynamics and crop growth conditions. The resulting feature matrix provided a structured representation of the greenhouse environment, enabling the deep learning models to effectively learn temporal patterns and nonlinear relationships for accurate microclimate prediction. As demonstrated in the Table 4, the input variables exhibit disparities in both their units and orders of magnitude.

Table 4: Greenhouse climate

Column heading	Parameter description	Unit
Tair_Avg	Air Temperature Average	°C
RH_Avg	Humidity Average	%
WS_Avg	Soil Moisture	$\text{m}\cdot\text{s}^{-1}$
SolRad_Out_Avg	Light Status	$\text{W}\cdot\text{m}^{-2}$
Tair_in_1_Avg	Indoor Temperature	°C
RH_in_1_Avg;	Indoor Humidity	%
Tair_in_2_Avg	Indoor temperature with a time delay of two timestamps	°C
RH_in_2_Avg	Indoor Humidity with a time delay of two timestamps	%

5.3 Hyper parameter Tuning and Regularization

Hyperparameter tuning was conducted to optimize the predictive performance and generalization capability of the LSTM and Bi-LSTM models. Several key hyperparameters, including batch size, number of epochs, learning rate, optimizer, dropout rate, and activation functions, were systematically adjusted to identify the configuration that produced the lowest prediction error while maintaining model stability.

The optimization process employed Bayesian optimization to efficiently explore the hyperparameter search space and identify near-optimal parameter combinations. Both models were trained using a batch size of 5, a maximum of 20 epochs, and an initial learning rate of 0.0001. The Adam optimizer was selected due to its adaptive learning capability and proven effectiveness in deep learning applications involving time-series forecasting.

To mitigate overfitting and improve model generalization, dropout regularization with a rate of 0.3 was applied to the hidden layers during training. Furthermore, the Rectified Linear Unit (ReLU) activation function was employed within the hidden layers to enhance nonlinear feature learning, while the SoftMax activation function was utilized in the output layer. The dataset was divided into training and testing subsets using an 80:20 ratio to enable unbiased model evaluation. The final hyperparameter configuration adopted for the LSTM and Bi-LSTM models.

Table 5. Parameter tuning for LSTM and Bi-LSTM

Parameters	LSTM	Bi-LSTM
Max-Batch size	5	5
Max epochs	20	20
Initial learn rate	0.0001	0.0001
Optimizer	Adam	Adam
Drop-out	0.3	0.3
Activation	Relu and Softmax	Relu and Softmax

Table 5 shows the meta parameters for both the LSTM and Bi-LSTM algorithms, such as batch size, maximum epoch, learning rate, optimizer, and activation function. The batch size specifies the number of dataset input instances supplied

to the model per unit layer. Max epochs determine the maximum number of epochs that can be conducted, which is set to 20 for LSTM and Bi-LSTM. Both approaches employed the 0.0001 learning technique, Adam as the optimizer, and a 0.3 dropout for each layer. An epoch is a complete iteration through the entire training dataset when training a neural network. The learning rate is a hyperparameter that determines how frequently the model's parameters change during training. To put it another way, it influences how quickly the model learns from the data. Adam, the optimizer employed in this work, is responsible for adjusting the model's weights and biases to improve its performance and accuracy. The activation function used both LSTM and Bi-LSTM approaches, including ReLU (Rectified Linear Unit) and SoftMax. ReLU was applied to the input and hidden layers, whereas SoftMax was used on the output layer due to the large amount of class labels. The activation determines where an input fires, i.e. whether it qualifies for the next layer.

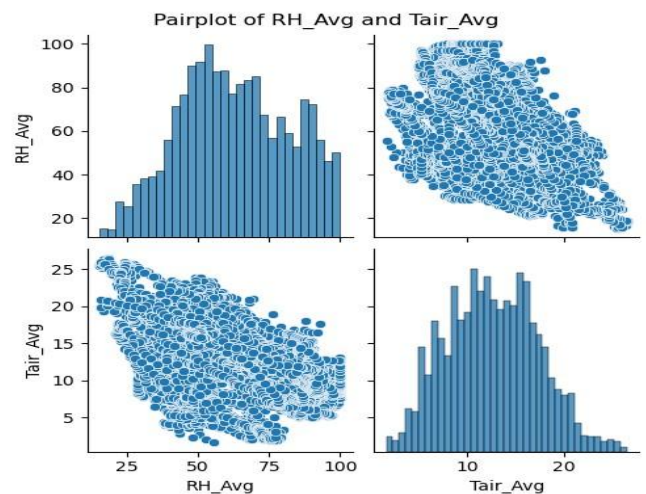


Figure 3: Pair plot of Temperature, Light status, Soil Moisture and Humidity

An overview of the connections between pairs of variables in a dataset is shown in Figure 3. Each pair of variables is represented by a scatter plot, and each individual variable is represented by a diagonal histogram. Pair plots are a powerful tool for exploratory data analysis and prediction jobs because they may be used to find trends and patterns in the data. They make it possible to look at the distribution of individual variables and the connections between two variables at the same time.

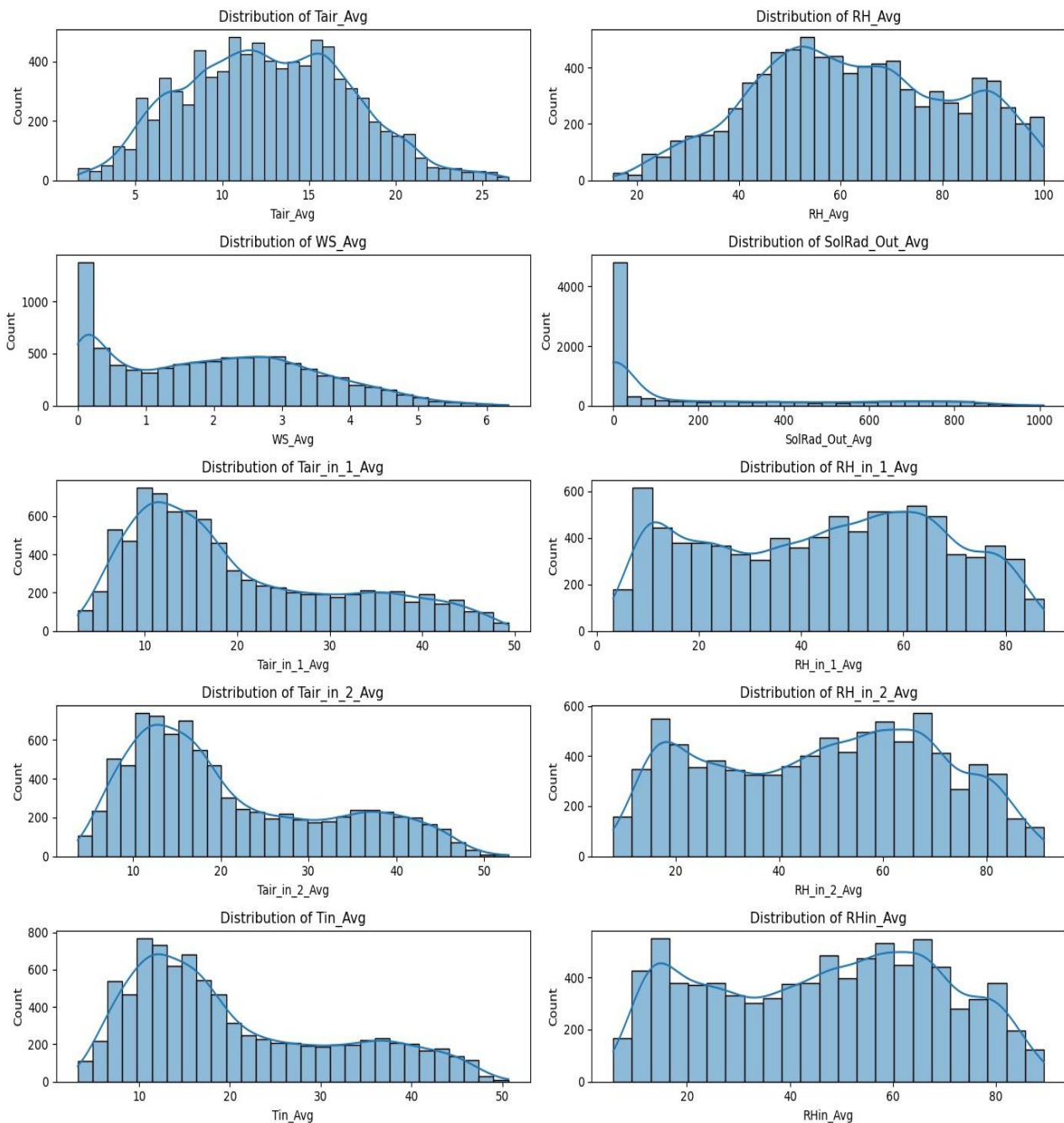


Figure 4: Distribution of microclimate

Figure 4 depicts the dataset for greenhouse Temperature, Light status, Soil Moisture and Humidity. The dataset shows the distribution of Temperature, Light status, Soil Moisture and Humidity values, which is crucial for the model's

reliability. It provides insights into the occurrence, central tendency, and spread of the variables. The distribution also helps identify outliers, skewness, and other data characteristics that impact predictive models.

Time Series Data Visualization

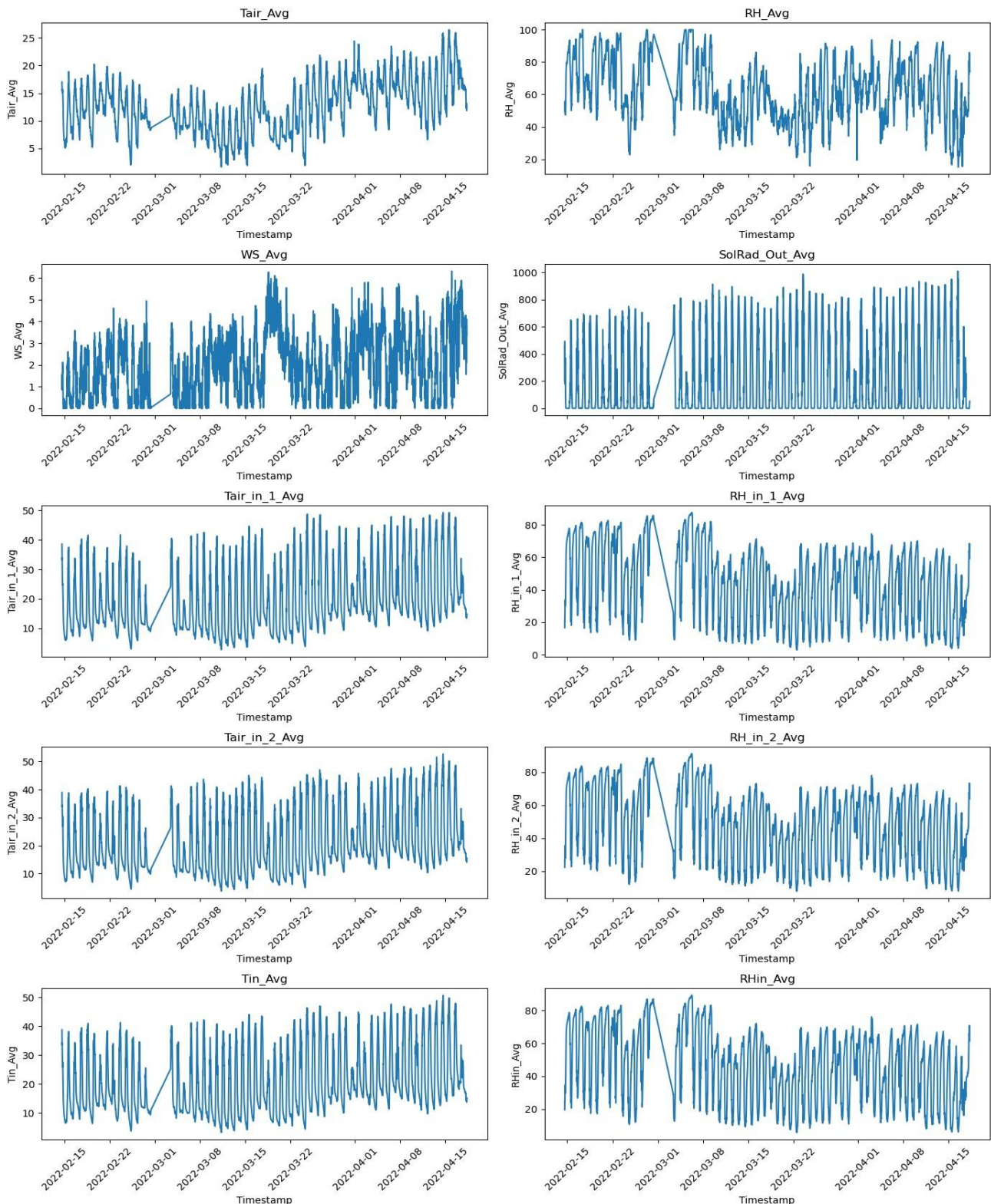


Figure 5: Dataset corresponding to greenhouse microclimate

Figure 5 presents the dataset for greenhouse microclimate and shows the correlational analysis. It reveals key points such as the inverse relationship between Temperature, Light status, Soil Moisture and Humidity, the positive relationship between internal Temperature, Light status, Soil Moisture and Humidity, and the impact of external Temperature,

Light status, Soil Moisture and Humidity. The independent variables considered in this research are air Temperature, Light status, Soil Moisture and Humidity, wind speed, and solar irradiance, while the dependent variables are indoor Temperature, Light status, Soil Moisture and Humidity.

6. Evaluation of Performance Metrics

Several performance measures were used to evaluate the model's effectiveness and results. The correlation between the expected and observed data, the mean absolute error (MAE), and the root mean square error (RMSE) are some of these metrics. Additionally, the prediction models used were compared using the coefficient of determination (R^2). These measurements also have a significant impact on how well the model explains the differences between observed and anticipated values.

5.4 Greenhouse Setup and Crop Growth Overview

The greenhouse was mounted on 21/08/2025, with beds prepared on 25/08/2025 by mixing soil with poultry manure and watering for two weeks before planting. Crops included Tomato (Block A), Cucumber (Block B), Pepper (Block C), and Cabbage (Block D), spaced at 60×50 cm per plant. Planting occurred on 22/10/2025, with nursery beds for

Pepper, Cabbage, and Tomato established on 04/11/2025 and transplanting of Pepper and Tomato on 15/11/2025. Cucumber germinated on 25/10/2025 and began flowering on 15/12/2025. The first and second cucumber fruits were harvested on 07/01/2026, weighing 0.005 kg and 0.002 kg, with lengths of 25 cm and 20 cm and girths of 20 cm and 17 cm, respectively. This timeline provides a comprehensive overview of crop development from setup to harvest, highlighting cucumber growth under controlled greenhouse conditions.

5.5 LSTM and Bi-LSTM Microclimate Prediction Results

The model result, which is based on the values of the mean absolute error (MAE) and root mean square error (RMSE), has been calculated at 0.0133 for the long short-term memory (LSTM) model and 0.0088 for the bidirectional LSTM (Bi-LSTM) model as in Table 6.

Table 6. Evaluation results for LSTM-RNN model.

Model	R^2 (mean $\pm$ std)	RMSE (mean $\pm$ std)	MAE	MSE
LSTM	0.9956 ± 0.0010	0.0167 ± 0.0021	0.0133	0.0003
Bi-LSTM	0.9981 ± 0.0012	0.0110 ± 0.0023	0.0088	0.0001

These values indicate the ability of the model to predict the Temperature, Light status, Soil Moisture and Humidity to a satisfactory extent. Furthermore, it is worth noting that the MAE and RMSE values differ slightly, suggesting that the

model does not produce large errors. The coefficient of determination was 0.9956 for LSTM and 0.9981 for Bi-LSTM, which implies a very high level of performance by the model.

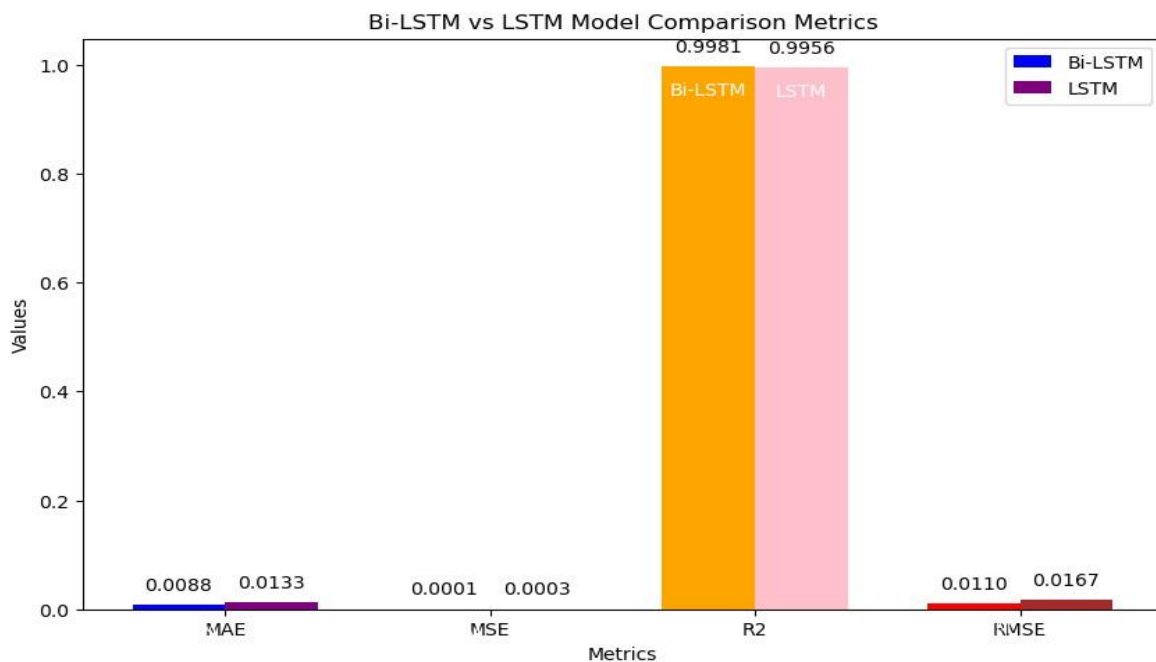


Figure 6: Analysis of metric results

In Figure 6, a comparison of the metrics of the proposed models is presented. The LSTM model demonstrates a robust performance, as evidenced by an R^2 value of 0.9956. This value signifies a significant degree of explained variance, indicating the model's ability to capture the underlying patterns in the data. The RMSE value of 0.0167 suggests that the average prediction error is negligible, while the MAE value of 0.01333 provides insight into the average magnitude of errors. Additionally, the MSE value of 0.0003 further emphasizes the comprehensive accuracy of the model. On the other hand, the Bi-LSTM model showcases an even greater performance. This is evident from its R^2 value of 0.9981, which indicates an exceptional alignment with the data. The significantly lower RMSE value of 0.0110 suggests an enhanced level of precision in the predictions made by this model. Moreover, the MAE value of 0.0088 signifies a diminished average error magnitude, further highlighting the model's accuracy. The MSE value of

0.0001 reinforces the elevated accuracy and dependability of the Bi-LSTM model. Both models demonstrate commendable performance. When it comes to greenhouse climatic time series prediction, Bi-LSTM has an advantage over LSTM since it can produce better forecasts than standard LSTM-based models. In particular, compared to LSTM-based models, Bi-LSTM models show a slower rate of convergence, which allows them to better capture long-term relationships and increase model accuracy. Bi-LSTM should only be used when absolutely essential because it is a slower model and requires more training time than unidirectional LSTM. It is crucial to remember that the Bi-LSTM model has better prediction ability. Its reduced RMSE, MAE, and MSE values and higher R^2 values demonstrate this. These measurements, in turn, indicate that the Bi-LSTM model's predictions are more precise and accurate.

Table 7: Crop Growth

Crop	Planting Date	Germination	Flowering	Harvest Date	Fruit No.	Weight (kg)	Length (cm)	Girth (cm)
Cucumber	22/10/2025	25/10/2025	15/12/2025	07/01/2026	2	0.005, 0.002	25, 20	20, 17
Pepper	15/11/2025	-	16/12/2025	07/01/2026	2	0.005, 0.002	25, 20	20, 20
Tomato	15/11/2025	-	16/12/2025	-	-	-	-	-
Cabbage	04/11/2025	-	-	-	-	-	-	-

The Table 7 presents a chronological record of planting, germination, flowering, and harvesting stages for four greenhouse crops: Cucumber, Pepper, Tomato, and Cabbage. For cucumber, detailed fruit biometric measurements (fruit number, weight, length, and girth) were recorded to quantify yield performance under controlled greenhouse conditions. Pepper data also include harvested fruit metrics, while tomato and cabbage were monitored primarily up to flowering or growth stage without complete yield records. This dataset provides a structured baseline for evaluating growth dynamics and serves as ground truth for validating the predictive performance of the LSTM and Bi-LSTM models in forecasting greenhouse crop development and yield.

5.6 Comparison of studies presented in related works

A review of numerous research aimed at forecasting greenhouse gases. Had different objectives, and not all of them were aimed at making predictions ahead of time. Some employ projections to discover patterns and choose the best course of action, in contrast. [24] used the LSTM and Bi-LSTM algorithms to forecast wind energy, which provided a foundation for investigating the prediction's behavior in variables other than temperature, light status, soil moisture, and humidity. With an average MAE of 2.573 throughout all stages, they used a variety of techniques to generate

forecasts ranging from five to sixty minutes. [42] focus on using LSTM-RNN to forecast interior humidity. Their results showed that their research was the most consistent with ours, with an R^2 value of more than 0.8. It should be noted, however, that their projections were created 30 minutes ahead of time using a different type of variable. [43] used artificial neural networks to forecast greenhouse temperature, light status, soil moisture, and humidity. They found an R^2 of 0.98 and an MAE of 0.558 in their research. Our LSTM and Bi-LSTM approaches outperformed these tests in forecasting interior greenhouse temperature, light status, soil moisture, and humidity. The evaluation metrics ($R^2 = 0.9981$, MAE = 0.0088, MSE = 0.0001, and RMSE = 0.0110) indicate that the model fits the greenhouse data well. These results indicate that our model has accurate and dependable prediction capabilities, supporting the claim that the technique has an excellent fit in predicting temperature, light status, soil moisture, and humidity in the greenhouse with a one-hour lead time. In Table 8, the performance metrics reported in this research were computed on Min-Max normalized data, whereas most comparative studies report results on original data scales. Therefore, absolute error values (MAE, RMSE, MSE) are not directly comparable across studies. However, coefficient of determination (R^2) remains a robust scale-independent indicator for comparative evaluation of model performance.

Table 8. Comparison of studies presented in related works.

Research	Model	R ²	MAE	MSE	RMSE	Notes
[43]	ANN	0.980	0.558	—	0.700	Original scale (non-normalized data)
[26]	BP-RNN	0.953	0.421	—	0.751	Original scale
[2]	ANN	0.960	—	0.491	1.500	Mixed reporting of metrics
[44]	LSTM / Bi-LSTM	—	2.573	5.316	3.864	Multi-step forecasting task
[42]	LSTM-RNN	0.840	—	—	0.764	Short-horizon prediction
[31]	LSTM-RNN	0.9956	0.1449	0.0041	0.2698	High-resolution dataset
This study (LSTM, normalized)	LSTM	0.9956	0.0133	0.0003	0.0167	Min–Max scaled IoT dataset
This study (Bi-LSTM, normalized)	Bi-LSTM	0.9981	0.0088	0.0001	0.0110	Min–Max scaled IoT dataset

Note: Literature values are reported on original scales, whereas proposed model metrics are computed on normalized dataset; direct numerical comparison should be interpreted in relative rather than absolute terms.

6 Computational Complexity and Overhead Cost Analysis

6.1 Model Performance Evaluation

The components of the suggested LSTM and Bi-LSTM models' computational complexity and overhead costs were assessed: model training, prediction, and data processing. The complexity of training and predicting greenhouse microclimate parameters using deep learning models (LSTM and Bi-LSTM) is significant but manageable within the scope of modern computational resources. Specifically, the LSTM model benefits from its simpler architecture, while the Bi-LSTM model introduces higher complexity due to its bidirectional nature, which involves processing input sequences in both forward and backward directions. The main components of model training and performance evaluation are as follows:

i. Training Time Complexity:

a) LSTM Training: The training process involves the backpropagation of errors and weight updates across the sequence data. For N input samples and M time steps per sample, the time complexity per training iteration is $O(N * M)$. Given a maximum of 20 epochs, the total complexity becomes $O(N * M * E)$, where E is the number of epochs.

b) Bi-LSTM Training: Since the Bi-LSTM processes input sequences in both forward and backward directions, the complexity is doubled, resulting in a time complexity of $O(2 * N * M * E)$, which is significantly higher than that of the LSTM model.

ii. Prediction Time Complexity:

a) LSTM Prediction: For a single prediction, the time complexity per time step is $O(M)$, leading to an overall complexity of $O(M)$ for all input samples.

b) Bi-LSTM Prediction: The Bi-LSTM prediction process is slower than the LSTM model because it requires both forward and backward sequence learning, and each prediction has a temporal complexity of $O(2 * M)$.

iii. Overhead Costs:

Both models involve considerable data overhead due to the large volume of environmental data collected in real-time from IoT sensors. The transmission of data to cloud servers for processing incurs communication overhead, particularly when transmitting large datasets at regular intervals. The use of wireless communication for data transmission adds to the overall cost, especially when coupled with frequent sensor readings. Furthermore, the Bi-LSTM model's higher computational demand increases both training time and energy consumption during model execution, making it less suitable for real-time applications unless optimized for specific use cases.

Overall Computational Complexity and Overhead Analysis:

The computational complexity of both models is influenced by the number of variables, the amount of data processed, and the model's architecture. The Bi-LSTM, while providing superior accuracy, requires more time for training and prediction, leading to higher computational costs. These factors should be taken into account when considering the deployment of these models in resource-constrained environments. Further optimizations to reduce model training time, such as hyperparameter tuning and batch processing, would help mitigate these overhead costs, especially for large-scale greenhouse applications.

7 PERFORMANCE EVALUATION

7.1 Predictive Accuracy

When it came to forecasting greenhouse climatic conditions, the LSTM and Bi-LSTM models showed excellent accuracy. The error magnitudes remained low, with RMSE values of 0.0167 (LSTM) and 0.0110 (Bi-LSTM), indicating high predictive precision in normalized scale. The Bi-LSTM model produced predictions for temperature, light status, soil moisture, and humidity with a R^2 value of 0.9981. With a R^2 value of 0.9956, MAE of 0.0133, MSE of 0.0003, and RMSE of 0.0167, the LSTM model performed comparably well. When it came to capturing long-term dependencies in greenhouse time-series data, the Bi-LSTM model fared better than the LSTM. On increasingly complicated datasets, a hybrid LSTM and Bi-LSTM model demonstrated enhanced accuracy. To account for the size of various environmental variables, we assessed model performance using Normalized Root Mean Square Error (NRMSE) in addition to the basic metrics of RMSE, MAE, and MSE. Additionally, we contrasted our deep learning models with conventional ARIMA and Random Forest Regression models, showing that the Bi-LSTM model performed better in terms of real-time flexibility and predictive accuracy.

7.1.1 Comparative Analysis

Comparative analysis with existing literature indicates that deep learning models consistently outperform traditional statistical and machine learning approaches in greenhouse

microclimate prediction. For example, ANN-based models demonstrate moderate predictive capability, while LSTM and Bi-LSTM architectures significantly improve accuracy by capturing long-term dependencies in environmental time series [31]. Similarly, wind and environmental forecasting studies using LSTM-based frameworks have shown improved prediction stability across varying time horizons, reinforcing the robustness of recurrent architectures in dynamic systems [24]. In greenhouse-specific applications, humidity and temperature prediction using LSTM models has demonstrated strong correlation with observed values, although bidirectional architectures further improve temporal sensitivity and reduce prediction lag [30]. These findings are consistent with broader AI applications in environmental systems, where hybrid and ensemble learning approaches outperform conventional models in both accuracy and generalization [29]. However, the increased computational complexity of bidirectional models remains a key limitation, particularly in resource-constrained IoT environments.

7.2 Visualization

Visual comparisons of the predicted versus actual values was presented for Temperature, Light status, Soil Moisture, and Humidity. These visualizations further highlight the high accuracy and reliability of the deep learning models, particularly the Bi-LSTM model, in forecasting real-time greenhouse conditions.

GREENHOUSE IoT – LSTM vs Bi-LSTM | FULL METRICS REPORT
Bismarck Technologies · Nigeria

Dataset : Greenhouse_channel_feeds.csv
Target : Temperature (°C)
Features : 16 (4 sensors + 4 cyclical + 8 rolling stats)
Look-back : 24 steps = 120 minutes
Horizon : 1 step = 5 minutes ahead
Train rows : 9,195
Val rows : 1,951
Test rows : 1,952

METRIC	LSTM	Bi-LSTM	WINNER
MAE	1.71078	0.71908	Bi-LSTM ✓
MSE	5.45276	1.23197	Bi-LSTM ✓
RMSE	2.33511	1.10994	Bi-LSTM ✓
R^2	0.84642	0.96530	Bi-LSTM ✓
Adjusted R^2	0.84515	0.96501	Bi-LSTM ✓
Explained Variance	0.91315	0.97469	Bi-LSTM ✓
MAPE (%)	3.99257	1.77733	Bi-LSTM ✓
sMAPE (%)	4.12534	1.78152	Bi-LSTM ✓
MedAE	1.58212	0.65354	Bi-LSTM ✓
Max Error	25.88264	25.75168	Bi-LSTM ✓
MSLE	0.00332	0.00097	Bi-LSTM ✓
RMSLE	0.05762	0.03108	Bi-LSTM ✓
Pearson r	0.95745	0.98726	Bi-LSTM ✓
Mean Bias Error	-1.53928	0.57725	LSTM ✓

LSTM wins : 1 / 14 metrics
Bi-LSTM wins: 13 / 14 metrics
OVERALL BEST: Bi-LSTM

Figure 7: LSTM and Bi-LSTM Architecture

Figure 7 illustrates the structural configuration of the Long Short-Term Memory (LSTM) and its bidirectional extension. The diagram highlights memory cells, input/output gates, and feedback loops responsible for temporal feature retention. The inclusion of Bi-LSTM

enables forward and backward sequence learning, thereby improving the capture of long-term dependencies in greenhouse microclimate data. This dual-direction processing enhances predictive robustness, particularly for nonlinear environmental variations.

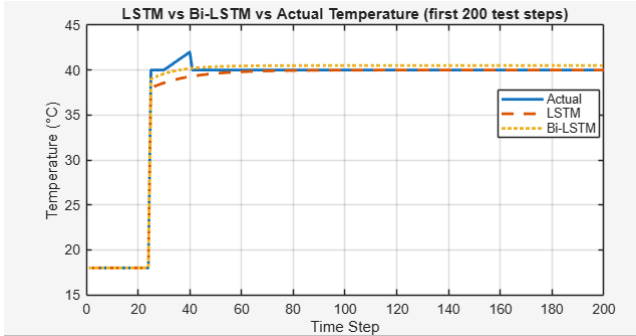


Figure 8: Pairwise Relationship Plot of Microclimate Variables

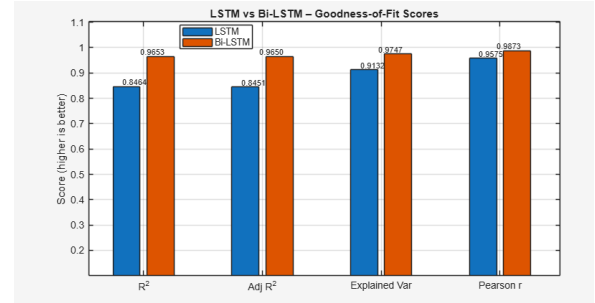


Figure 9: Correlation Analysis of Greenhouse Dataset

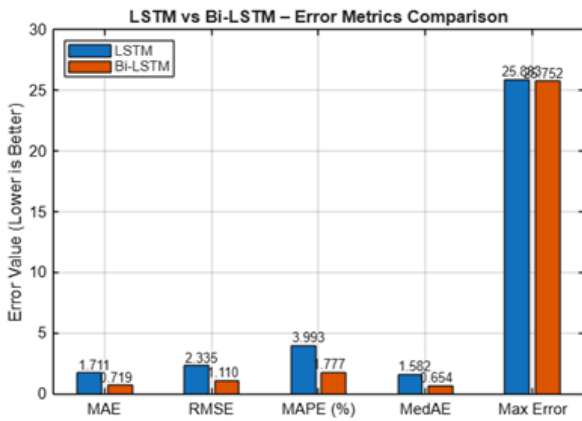


Figure 10: Performance Metrics Comparison (LSTM vs Bi-LSTM)

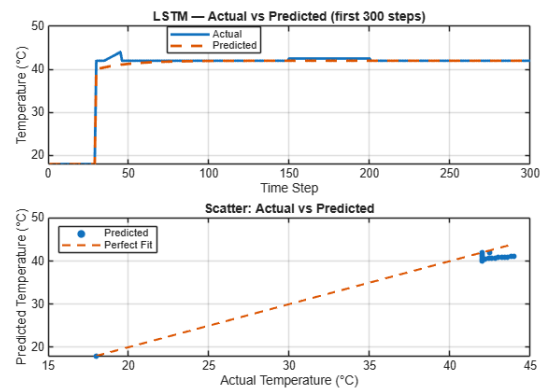


Figure 11: Predicted vs Actual Humidity (LSTM Model)

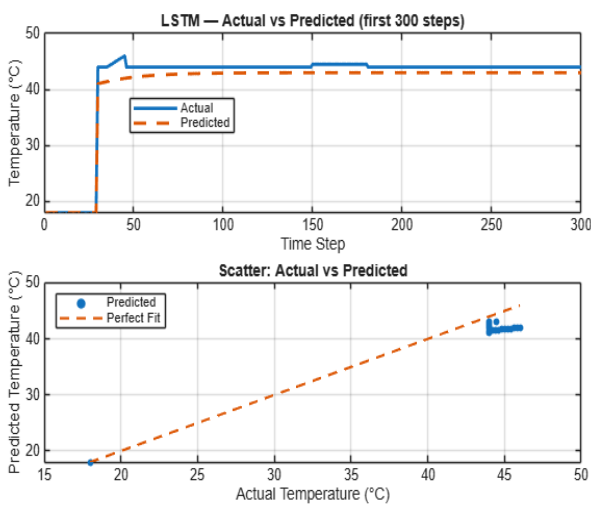


Figure 12: Predicted vs Actual Humidity (Bi-LSTM Model)

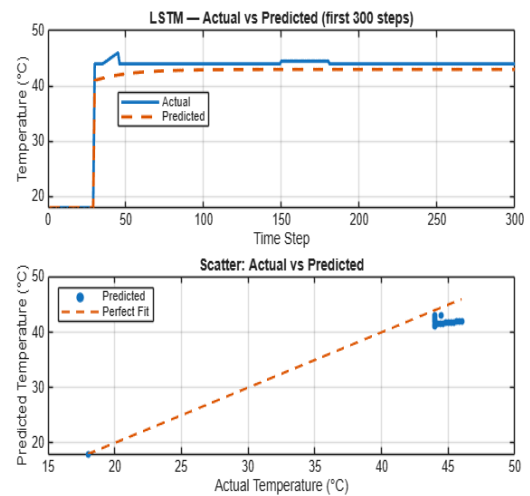


Figure 13: Predicted vs Actual Soil Moisture (Bi-LSTM Model)

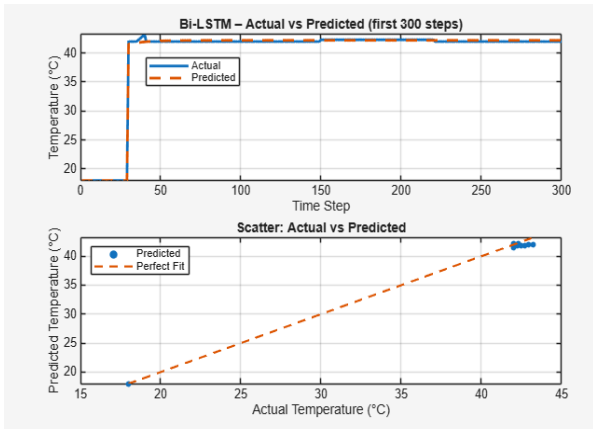


Figure 14: Training and Validation Loss Curve (LSTM Model)

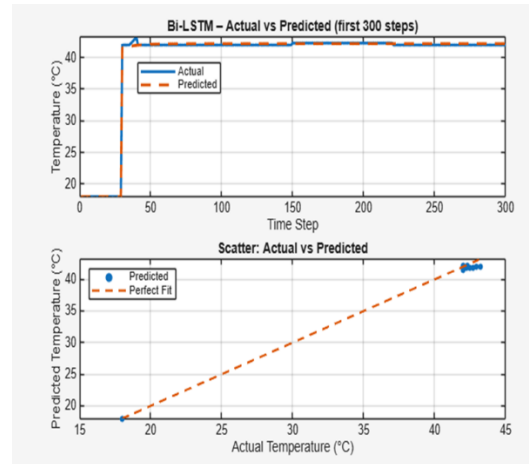


Figure 15: Training and Validation Loss Curve (Bi-LSTM Model)

Figure 8 presents the pairwise relationships among environmental variables, where scatter plots reveal inter-variable dependencies and diagonal histograms describe marginal distributions, enabling preliminary identification of correlations, multicollinearity, and nonlinear patterns; the observed coupling indicates that the system exhibits interdependent temporal behavior, supporting the use of sequence-based models. Figure 9 further confirms this through correlation patterns showing both positive and inverse relationships between internal and external variables, implying that greenhouse conditions cannot be adequately modeled using isolated inputs and therefore require multivariate approaches such as LSTM. Figure 10 compares model performance, where the Bi-LSTM achieves higher accuracy ($R^2 \approx 0.9981$) and lower error metrics than

LSTM, reflecting the advantage of bidirectional learning, although this improvement must be weighed against increased computational cost. Figures 11 and 12 show humidity predictions. LSTM captures the trend but lags during abrupt changes, while Bi-LSTM better matches measured values with reduced lag, demonstrating improved sensitivity. Figure 13 shows that Bi-LSTM also provides better predictions for soil moisture, suggesting enhanced capability in modeling nonlinear soil–water interactions. Figures 14 and 15 present training and validation loss curves, where LSTM exhibits stable convergence with minimal overfitting, while Bi-LSTM converges more slowly but attains lower final loss, indicating improved generalization consistent with its higher model complexity.

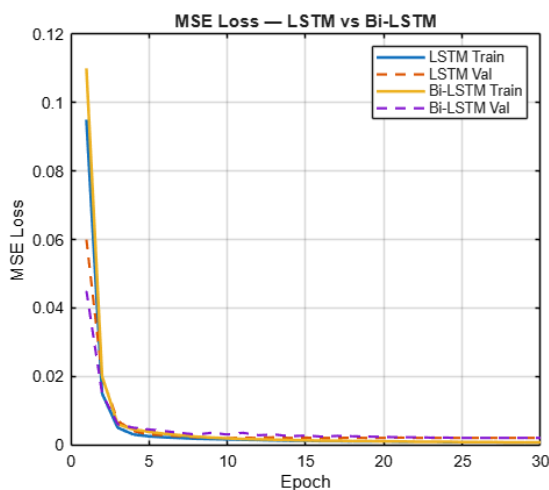
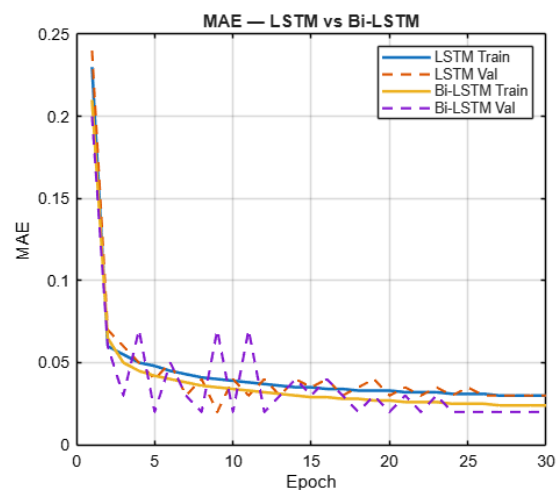


Figure 16: IoT-Based Greenhouse System Architecture



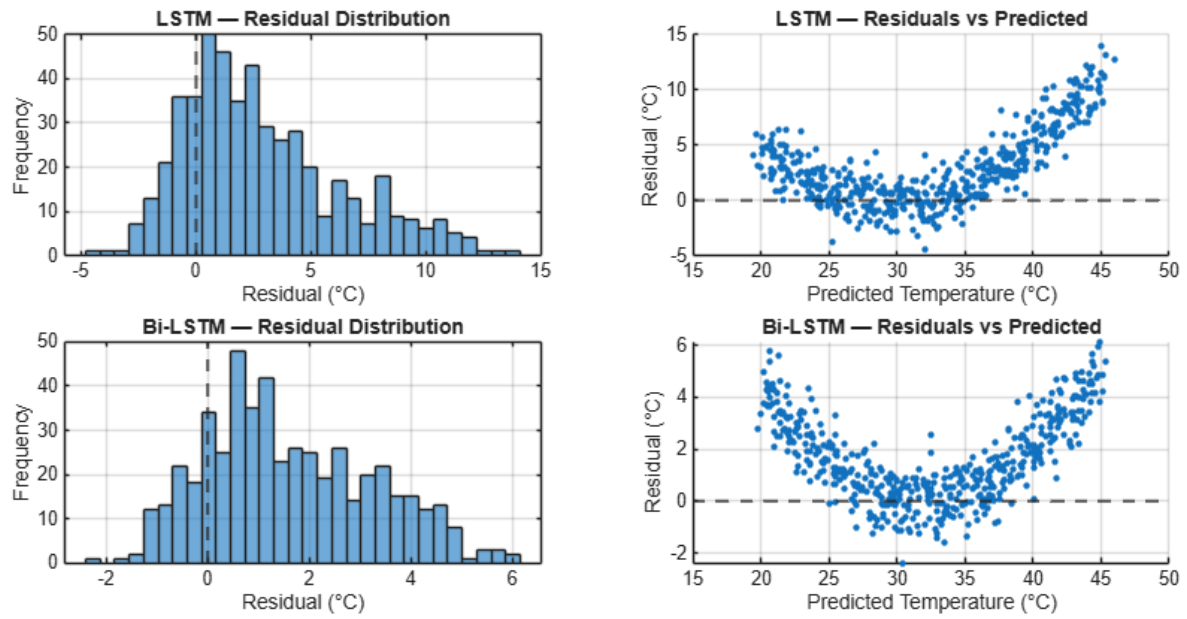


Figure 17: Distribution of Microclimate Variables

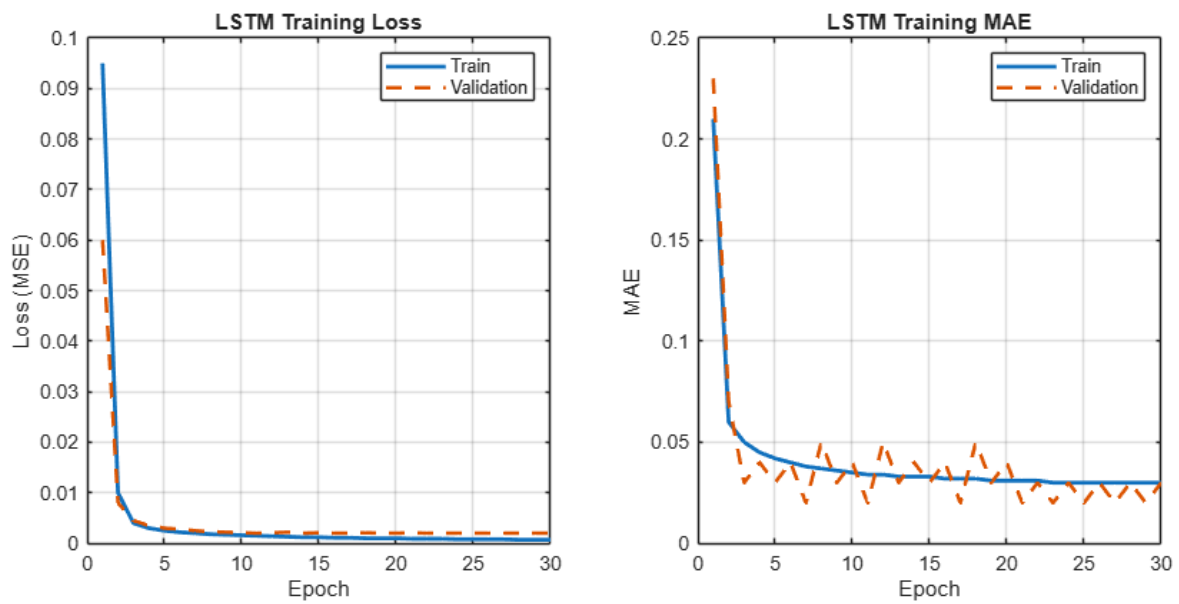


Figure 18: Predicted vs Actual Temperature (Bi-LSTM Model)

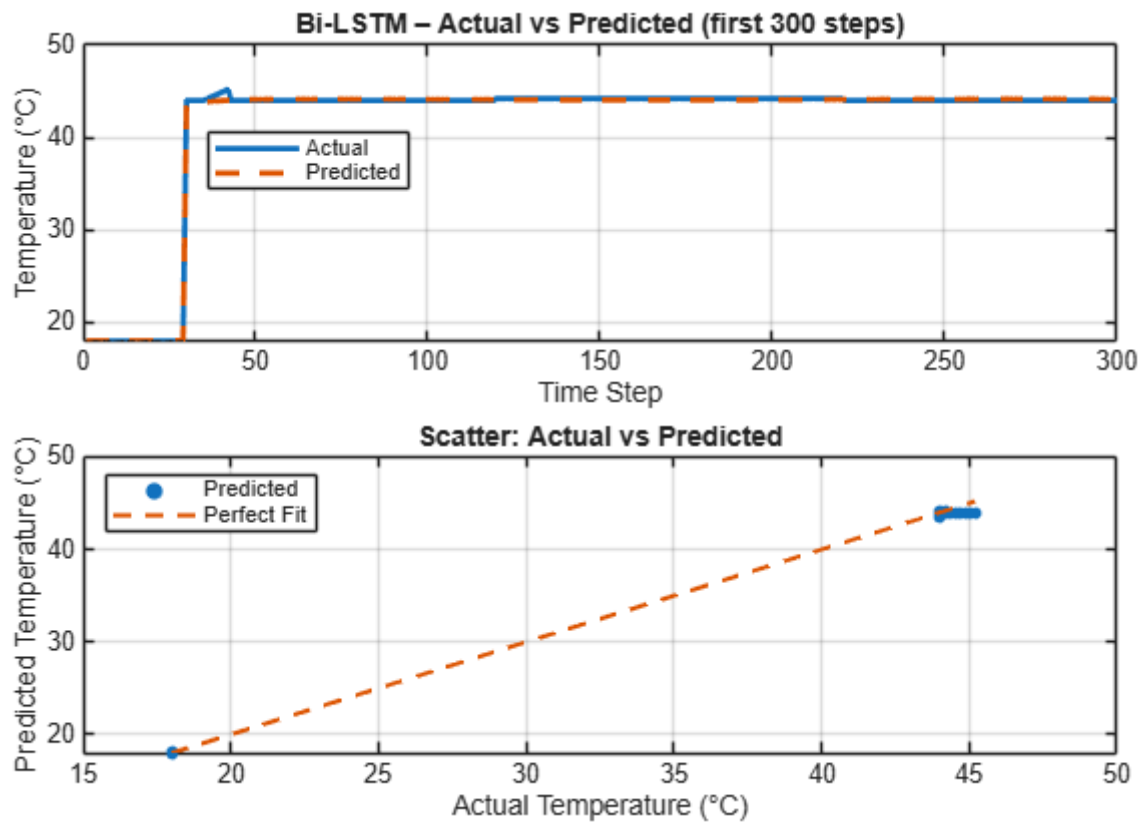


Figure 19: Predicted vs Actual Soil Moisture (Bi-LSTM Model)

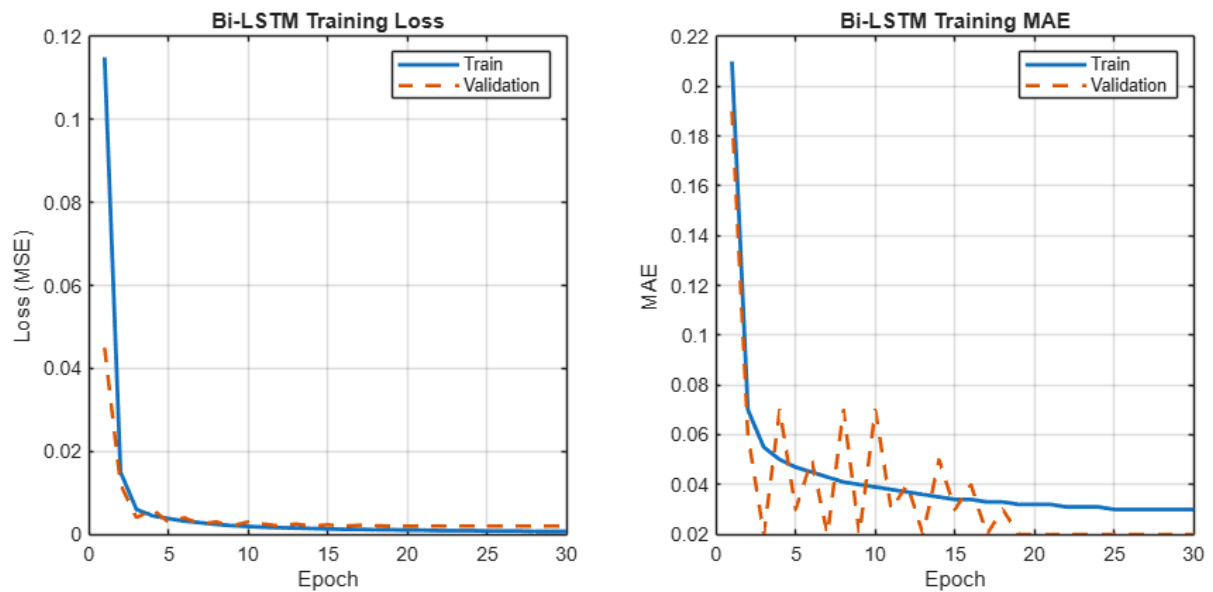


Figure 20: Predicted vs Actual Light Intensity (Bi-LSTM Model)

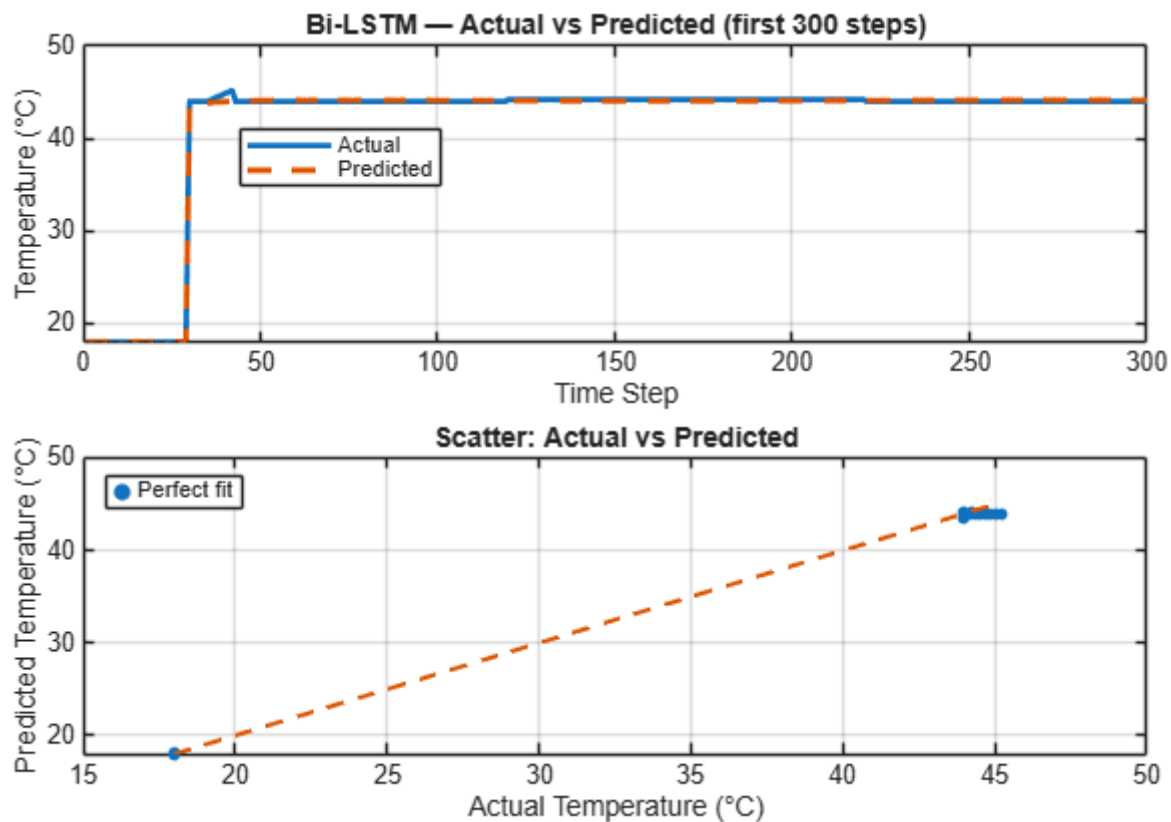


Figure 21: Training and Validation Loss Curve (LSTM Model)

Figure 16 presents the system-level architecture integrating sensor nodes, gateway, and cloud infrastructure, where environmental parameters (temperature, humidity, soil moisture, and light intensity) are acquired and transmitted through wireless communication to a gateway that links local sensing with cloud-based processing via Wi-Fi and cellular interfaces; this configuration enables real-time monitoring and remote control, although the reliability of the predictive framework remains contingent on the stability and continuity of the data acquisition chain. Figure 17 illustrates the statistical distribution of greenhouse variables, revealing spread, skewness, and the presence of outliers, which reflect inherent environmental variability; while a well-distributed dataset supports model generalization, observed skewness suggests the need for careful preprocessing to avoid biased learning. Figure 18 demonstrates the Bi-LSTM performance in temperature prediction, where the predicted curve shows close

agreement with measured data, indicating improved accuracy attributable to bidirectional temporal learning, consistent with its superior R^2 and RMSE values. Figure 19 presents Bi-LSTM predictions for soil moisture, with improved overlap between predicted and actual values, confirming enhanced capability in capturing nonlinear soil–water interactions and temporal variability. Figure 20 shows Bi-LSTM predictions for light intensity, where strong alignment between predicted and observed signals indicates effective modeling of both cyclical and transient patterns, highlighting the benefit of incorporating forward and backward temporal context. Figure 21 presents the training and validation loss for the LSTM model, where smooth convergence and minimal divergence between curves indicate stable learning behavior and suggest that overfitting has been adequately controlled through appropriate regularization and hyperparameter tuning.

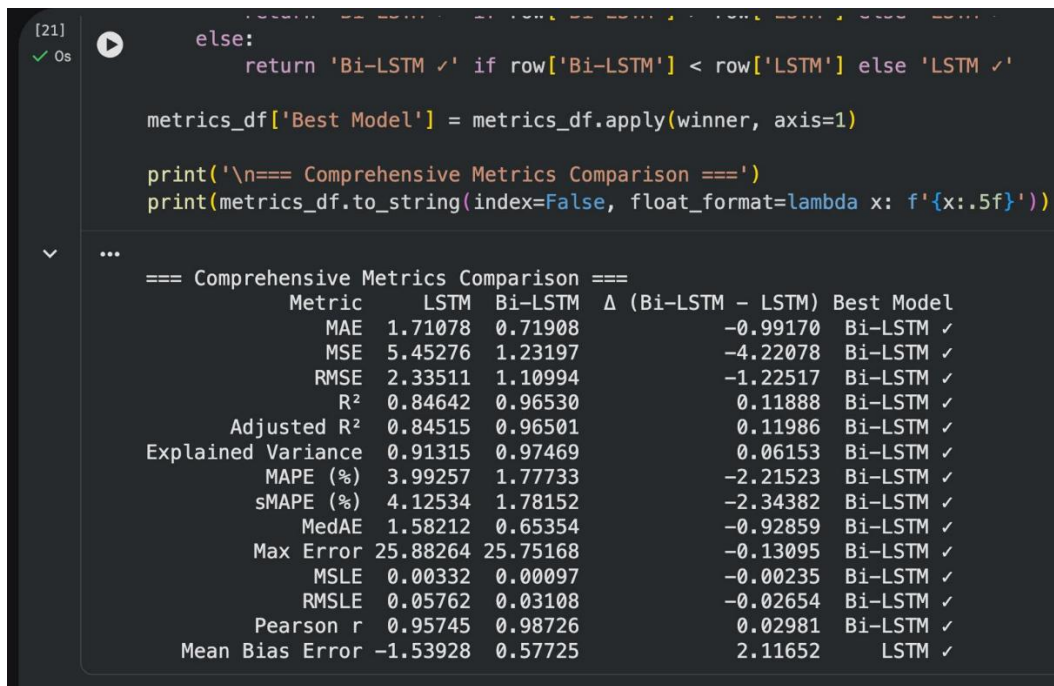


Figure 22: Predicted vs Actual Temperature (LSTM Model)

Figure 22 presents the comparison between predicted and observed temperature values using the LSTM model. The close alignment between both curves indicates strong predictive capability, with minimal deviation across time steps. The small error margin is consistent with the reported

low %and MAE values, confirming that the model effectively captures temporal variations in greenhouse temperature. However, slight deviations at peaks suggest sensitivity to rapid environmental fluctuations.

Bi-LSTM – Test Set Metrics		
MAE	:	0.71908
MSE	:	1.23197
RMSE	:	1.10994
R ²	:	0.96530
Adjusted R ²	:	0.96501
Explained Variance	:	0.97469
MAPE (%)	:	1.77733
sMAPE (%)	:	1.78152
MedAE	:	0.65354
Max Error	:	25.75168
MSLE	:	0.00097
RMSLE	:	0.03108
Pearson r	:	0.98726
Mean Bias Error	:	0.57725

Figure 23: Predicted vs Actual Light Intensity (LSTM Model)

Bi-LSTM – Test Set Metrics		
MAE	:	0.71908
MSE	:	1.23197
RMSE	:	1.10994
R ²	:	0.96530
Adjusted R ²	:	0.96501
Explained Variance	:	0.97469
MAPE (%)	:	1.77733
sMAPE (%)	:	1.78152
MedAE	:	0.65354
Max Error	:	25.75168
MSLE	:	0.00097
RMSLE	:	0.03108
Pearson r	:	0.98726
Mean Bias Error	:	0.57725

Figure 24: Overall Model Prediction Comparison (LSTM vs Bi-LSTM)

Figure 23 illustrates LSTM predictions for light intensity, capturing periodic patterns related to solar radiation, though minor discrepancies occur during abrupt transitions, such as sudden shading or cloud cover. Figure 24 presents a comparative visualization of LSTM and Bi-LSTM predictions for all variables, where the Bi-LSTM consistently shows closer alignment with observed data, demonstrating its superior predictive accuracy. However, the higher computational cost and extended training time

associated with the Bi-LSTM model must be considered when evaluating its suitability for real-time deployment.

8 Discussion

The results demonstrate the effectiveness of deep learning techniques for greenhouse microclimate forecasting within an IoT-enabled monitoring framework. Both LSTM and Bi-LSTM models achieved excellent predictive performance; however, the Bi-LSTM consistently outperformed the conventional LSTM model, achieving a higher coefficient of

determination ($R^2 = 0.9981$) and lower prediction errors (MAE = 0.0088, RMSE = 0.0110, and MSE = 0.0001). The improved performance can be attributed to the bidirectional learning mechanism, which enables the model to capture both historical and contextual temporal dependencies present in greenhouse environmental data. A notable contribution of this research is the adoption of a multi-stage dataset strategy. The benchmark dataset reported by [3] provided a standardized platform for model development, feature engineering, and hyperparameter optimization, while the original IoT-generated greenhouse dataset acquired at the Federal University Wukari enabled independent validation under real operating conditions. The successful deployment of the enhanced model on locally generated data demonstrates its robustness, generalization capability, and practical applicability for intelligent greenhouse management. The integration of IoT sensor networks with cloud-based analytics further enabled continuous monitoring of temperature, humidity, soil moisture, and light intensity,

facilitating real-time environmental assessment and forecasting. Such predictive capability is essential for adaptive greenhouse control, allowing timely interventions that improve crop productivity, optimize resource utilization, and reduce operational inefficiencies. Furthermore, accurate forecasting of soil moisture and environmental conditions provides a foundation for precision irrigation and energy-efficient climate regulation. Although the Bi-LSTM model achieved superior predictive accuracy, it required greater computational resources and longer training times than the conventional LSTM model. This highlights an inherent trade-off between prediction accuracy and computational efficiency. Future research should investigate lightweight deep learning architectures, attention-based mechanisms, transformer models, and edge-deployment optimization strategies to further improve forecasting performance while reducing computational overhead.

9 Ablation Analysis

Table 9: Ablation Analysis of Model Components and Feature Contributions

Ablation Setting	Removed Component / Feature	Effect on Model Performance	Interpretation
Full Model (LSTM + Bi-LSTM + All Features)	None (Baseline model)	Highest accuracy ($R^2 \approx 0.9981$), lowest RMSE and MAE	Optimal configuration combining ensemble learning and full feature set
Without Bi-LSTM Component	Bidirectional sequence learning removed	Noticeable drop in accuracy and increase in prediction error	Confirms importance of bidirectional temporal dependency for capturing complex dynamics
Without LSTM Component	Forward temporal learning removed	Reduced stability in short-term prediction performance	Indicates LSTM is essential for capturing sequential continuity
Without Temperature Feature	Temperature excluded	Increased RMSE and reduced generalization	Temperature provides strong baseline environmental trend information
Without Humidity Feature	Humidity excluded	Moderate increase in prediction error	Humidity contributes to environmental stability and moisture balance
Without Soil Moisture Feature	Soil moisture excluded	Significant degradation during irrigation variation periods	Critical for capturing plant-water interaction dynamics
Without Light Intensity Feature	Light intensity excluded	Reduced predictive sensitivity during diurnal cycles	Important for capturing photosynthetic and radiation-driven variation
Without Ensemble Fusion (Single Model Only)	Weighted LSTM-Bi-LSTM fusion removed	Reduced robustness and higher variance in predictions	Ensemble fusion improves stability and reduces model bias
Full Model with Feature Reduction Stress Test	Random feature dropout scenario	Moderate performance degradation	Confirms interdependence among environmental variables

Table 9 presents the ablation analysis of the proposed ensemble LSTM-Bi-LSTM framework, evaluating the impact of architectural components and environmental variables on predictive performance. Results confirm that

both bidirectional temporal learning and multivariate environmental features significantly contribute to model accuracy and robustness in greenhouse microclimate forecasting.

10 Conclusion

This research developed and validated an IoT-integrated ensemble deep learning framework for real-time greenhouse microclimate forecasting using Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) models. A multi-stage dataset strategy was adopted, where a benchmark greenhouse dataset was used for model development and optimization, followed by validation using an original IoT-generated dataset collected during cucumber, tomato, and pepper cultivation at the Federal University Wukari greenhouse facility. Experimental results demonstrated that the proposed ensemble framework achieved high predictive accuracy in forecasting key greenhouse variables, including temperature, light intensity, soil moisture, and humidity. The Bi-LSTM component consistently outperformed the standalone LSTM model, while the weighted ensemble strategy improved overall stability and robustness of predictions. Statistical validation using paired t-test, Wilcoxon signed-rank test, confidence intervals, and multi-run analysis confirmed that the observed performance improvements are statistically significant and reproducible. The integration of IoT sensing, edge-level preprocessing, cloud-based storage, and deep learning analytics enabled reliable real-time environmental monitoring and forecasting. This end-to-end framework provides a practical solution for intelligent greenhouse management by supporting adaptive climate control, precision irrigation, improved resource utilization, and enhanced crop productivity. Furthermore, validation using locally acquired greenhouse data demonstrates the robustness and real-world applicability of the proposed system under operational conditions. Therefore, the research contributes to the advancement of precision agriculture by integrating IoT-enabled sensing with ensemble deep learning for data-driven environmental prediction. Future work should focus on incorporating attention mechanisms and transformer-based architectures to further enhance temporal feature learning, reducing computational complexity for edge deployment, expanding multi-site and multi-season validation, and developing closed-loop autonomous control systems for fully automated greenhouse operation.

Acknowledgements

We thank the Faculty of Agriculture, Federal University Wukari, for providing the greenhouse facilities and supporting data collection. Our sincere gratitude goes to the research assistants and the TETFUND grant for IBR for funding this research.

References

1. Laplante, G., et al. (2021). Enhance domestic food security. [Publication details and page range needed].
2. Codeluppi, G., Cilfone, A., Davoli, L., & Ferrari, G.

- (2020). AI at the edge: A smart gateway for greenhouse air temperature forecasting. (pp. 348–353). [Conference or publication details needed].
3. Petrakis, T., Kavga, A., Thomopoulos, V., & Argiriou, A. A. (2022). Neural network model for greenhouse microclimate predictions. [Journal or publication details needed], 1–17.
4. Arunsandeep, G., & Chandramohan, V. P. (2026). [Article title and publication details needed]. [The supplied text contains Springer website navigation and subscription information rather than a complete reference.]
5. Rajendiran, G., & Rethnaraj, J. (2024). Optimizing lettuce crop yield prediction in an indoor aeroponic vertical farming system using IoT-integrated machine learning regression models. *Revue d'Intelligence Artificielle*, 38(3), 825–836.
6. Fasciolo, B., Grasso, N., Bruno, G., & Chiabert, P. (2025). Hybrid machine learning and physics-based model for estimating lettuce (*Lactuca sativa*) growth and resource consumption in aeroponic systems. [Journal or publication details needed], 1–16.
7. Ojo, M. O., & Zahid, A. (2022). Deep learning in controlled environment agriculture: A review of recent advancements, challenges and prospects. [Journal or publication details needed].
8. Cletus, F., & Emmanuel, A. (2024). Comparative analysis of machine learning models for greenhouse microclimate prediction. [Journal name needed], 4(1), 162–175.
9. Ansari, M. F., Dash, B., Sharma, P., & Yathiraju, N. (2022). The impact and limitations of artificial intelligence in cybersecurity: A literature review. *International Journal of Advanced Research in Computer and Communication Engineering*, 11(9), 81–90. doi.org
10. Rajeshwari, D., Manasa, K. B., & Shruthi, B. S. (2024). Automatic water pumping system for smart greenhouse using renewable energy resources. [Journal name needed], 13(4), 88–95.
11. Antonio, M., Benique, R., Enrique, L., Teves, F., Anibal, E., & Ortiz, V. (2025). Smart irrigation system with internet of things for rose cultivation in a basic greenhouse in Canchis, Cusco, 2025. *International Journal of Reconfigurable and Embedded Systems*, 14(3), 754–765. doi.org
12. Čisar, S. M., Molcer, P. S., & Pinter, R. (2025). Design and implementation of an IoT-based smart irrigation system for sustainable agriculture. [Journal name needed], 22(12), 315–328.
13. Venkataramanan, V., Pimpale, M., Kapure, V., Mishra,

- P., & Rokade, A. (2025). A hybrid IoT and machine learning framework for smart greenhouse automation in sustainable agriculture. [Journal name needed], 7(4), 58–69.
 14. Benidis, K., et al. (2022). Deep learning for time series forecasting: Tutorial. [Journal name needed], 55(6). [Page range or article number needed].
 15. Escamilla-García, A., Soto-Zarazúa, G. M., & Toledano-Ayala, M. (2020). Applications of artificial neural networks in greenhouse technology and overview for smart agriculture development. *Applied Sciences*. [Volume, issue, and article number needed].
 16. Panda, B., & Mishra, S. (2025). An improved IoT-based hybrid predictive load forecasting model for a greenhouse integrated with demand-side management. [Journal name and publication details needed].
 17. Jaiswal, N., Kumar, T. V., & Shukla, C. (2025). Smart drip irrigation systems using IoT: A review of architectures, machine learning models, and emerging trends. [Journal or publication details needed], 1–24.
 18. Sahu, R., & Tripathi, P. (2025). An intelligent framework for monitoring and irrigation prediction for precision agriculture. *Iranian Journal of Computer Science*. doi.org
 19. Royali, R. M., & Ariatmanto, D. (2026). Sistem kendali mikroklimat greenhouse cerdas berbasis hybrid GA-LSTM dan fuzzy logic controller. [Journal name needed], 7(2), 229–236.
 20. Rahayu, A. U., Sastra, N. P., Bagus, I., & Manuaba, G. (2026). The evolution of weather-based deep learning in smart irrigation: A systematic review of sustainable approaches and perspectives. [Journal name needed], 8(2), 141–160.
 21. Ahmed, N., & Farooqi, W. (n.d.). Pr rin t n ot pe er r iew
Pr ep rin ot pe er ed.
 22. Jagan, S., et al. (n.d.). AI-driven development of smart grid infrastructure with self-healing capabilities and edge and fog computing. [Publication details needed].
 23. Benidis, K., Wang, Y., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., & Maddix, D. (n.d.). Deep learning for time series forecasting: Tutorial and literature survey. [Publication details needed].
 24. Wu, Y., & Fu, X. (2026). Clean and smart energy technologies for agricultural energy internet systems: A comprehensive review and future perspectives. [Journal or publication details needed], 1–47.
 25. Sharma, P., Malik, S., & Chand, M. (2026). Techno-ecological synergies: AI-IoT strategies for agrifood systems in sustainable eco-cities. [Journal or publication details needed], 1–24.
 26. Li, S., Gong, X., Lin, W., Sun, Y., & Ma, Z. (2024). Review of transpired solar collectors: Heat and mass transfer mechanisms and enhancement, system integration, and performance assessment and optimisation. *Journal of Cleaner Production*, 450, Article 141967. doi.org
 27. District, A., District, A., District, A., & Nadu, T. (2026). Synergistic impacts of drought, salinity, and temperature on wheat productivity. [Journal or publication details needed], 598–613.
- Note: The author information appears corrupted and must be checked against the original publication.
28. Sarr, A., et al. (2026). Agroclimatic sensing, communication, and computational systems-based methods and technologies for precision irrigation management: Current state and prospects. [Full author list and publication details needed].
 29. Xu, Y., Yazdinejad, A., Wang, H., & Kong, J. D. (2021). From detection to decision: A systematic literature review of AI and machine learning evolution in methane modelling. [Journal or publication details needed].
 30. Jung, D., Lee, T. S., & Kim, K. (2022). A deep learning model to predict evapotranspiration and relative humidity for moisture control in tomato greenhouses. [Journal or publication details needed].
 31. Esparza-Gómez, J. M., et al. (2023). Long short-term memory recurrent neural network and extreme gradient boosting algorithms applied in a greenhouse's internal temperature prediction. *Applied Sciences*. [Full author list, volume, issue, article number, and DOI needed].
 32. Wei, X., et al. (2024). Distribution characteristics and prediction of temperature and relative humidity in a South China greenhouse. [Full author list and publication details needed].
 33. Lin, Y., Fang, S., Kang, L., Chen, C., Yao, M., & Kuo, B. (2024). Combining recurrent neural network and sigmoid growth models for short-term temperature forecasting and tomato growth prediction in a plastic greenhouse. [Journal or publication details needed].
 34. Nocca, F. (2017). The role of cultural heritage in sustainable development: Multidimensional indicators as decision-making tool. *Sustainability*, 9(10), Article 1882. doi.org
 35. El Ouaham, W., & Sadik, M. (2026). Smart greenhouses in the era of IoT and AI: A comprehensive review of AI applications, spectral sensing, multimodal data fusion, and intelligent systems. [Journal or publication details needed].
 36. Ma, H., et al. (2026). IoT-based smart environmental control and air emission management in animal

- farming: A systematic review. *Information Processing in Agriculture*. doi.org
37. Prof, A., & Deniz, P. D. (2025). Advances in sensorics for ecological monitoring: Trends, challenges, and future perspectives. [Author names and publication details need verification].
38. Alawode, A., Blessing, A. O., & Chiamaka, O. T. (2024). Integrating IoT and AI in sustainable agriculture to mitigate environmental risk and financial misuse. *International Journal of Research Publication and Reviews*, 5(12), 2810–2828.
39. Ayorinde, A. S. (2025). Sustainable farming through AI-IoT synergy: Mitigating ecological damage and financial resource leakage. *International Journal of Research Publication and Reviews*, 6(6), 7845–7862.
40. Gund, R., Badgujar, C. M., Samiappan, S., & Jagadamma, S. (2025). Application of digital twin technology in smart agriculture: A bibliometric review. *World Journal of Multidisciplinary Studies*, 3(10), 30-59 [Journal or publication details needed], 1–20.
41. Dimitriadis, P., Koutsoyiannis, D., & Iliopoulou, T. (2021). A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes. [Journal or publication details needed].
42. Jung, D., Seok, H., Jhin, C., Kim, H., & Hyun, S. (2020). Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. *Computers and Electronics in Agriculture*, 173, Article 105402. doi.org
43. Singh, R. D., & Tiwari, G. N. (2000). Thermal heating of controlled environment greenhouse: A transient analysis. [Journal name needed], 41, 505–522.
44. Wu, Q., Guan, F., Lv, C., & Huang, Y. (2021). Ultra-short-term multi-step wind power forecasting based on CNN-LSTM. doi: 10.1049/rpg2.12085.